A Search for Neutral Higgs Bosons at High $\tan\beta$ in Multi–jet Events from $p\overline{p}$ Collisions at $\sqrt{s} = 1960$ GeV

Andrew C. Haas

A dissertation
submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

2004

Program Authorized to Offer Degree: Physics

University of Washington

Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Andrew C. Haas

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of Supervisory Committee:

_____

Gordon Watts

Reading Committee:

_____

Gordon Watts

_____

Toby Burnett

_____

Ann Nelson

Date: _____

University of Washington

Abstract

A Search for Neutral Higgs Bosons at High $\tan\beta$ in Multi–jet Events
from $p\overline{p}$ Collisions at $\sqrt{s} = 1960$ GeV

by Andrew C. Haas

Chair of Supervisory Committee:

Professor Gordon Watts
Physics

The Higgs mechanism preserves the gauge symmetries of the Standard Model while giving masses to the W, Z bosons. Supersymmetry, which protects the Higgs boson mass scale from quantum corrections, predicts at least 5 Higgs bosons, none of which has been directly observed. This thesis presents a search for neutral Higgs bosons, produced in association with bottom quarks. The production rate is greatly enhanced at large values of the Supersymmetric parameter $\tan\beta$. High–energy $p\overline{p}$ collision data, collected from Run II of the Fermilab Tevatron using the DØ detector, are analyzed. In the absence of a signal, values of $\tan\beta{>}80$–120 are excluded at 95% Confidence Level (C.L.), depending on the (CP–odd) neutral Higgs boson mass (studied from 100 to 150 GeV/c^2).

TABLE OF CONTENTS

ii

LIST OF FIGURES

viii

x

xiii

LIST OF TABLES

ACKNOWLEDGMENTS

To Mom and Dad, for their eternal love and support. Thanks to my advisor, Gordon Watts, for always being on my side. Thanks to all my advisors through the years: Dave Cutts, Toby Burnett, and Avto Kharchilava. Thanks to everyone at DØ and UW for making this work possible. Thanks to all my friends for keeping me going. And thanks to Ellen and Leon for entrusting us the Compound. I could not have done this without all of you.

Chapter 1

INTRODUCTION

The world we see around us is mostly made of atoms. Experiments in the early part of the last century, such as Rutherford's famous scattering experiment of alpha particles off gold foil, determined that atoms are composite structures containing a nucleus surrounded by a cloud of electrons. Later experiments showed that nuclei are made of even smaller particles, protons and neutrons. High energy particle beams showed that the protons and neutrons are made of still smaller point–like particles, quarks. At high energy accelerators, new particles were discovered which could not be explained without new, heavier quarks. Two particles with properties similar to the electron, but heavier, were also discovered: the muon (μ) and the tau (τ). Three very light, weakly interacting particles called neutrinos have also been directly observed [1].

The last century witnessed the birth of a rigorous theory, the Standard Model, describing all the observed particles and their interactions in terms of quantum fields. All particles seen in the laboratory are described by a field of the Standard Model [2].

However, the Standard Model requires at least one more field to explain the experimental data, which has not yet been observed directly: the *Higgs boson* field. The heaviest known quark, the top quark, was also predicted for consistency reasons, and was discovered at Fermilab in 1995. Symmetries of the Standard Model fields' interactions prevent them from directly acquiring masses. That masses are ob-

---

[1] There is a flavor of neutrino for each of the "electron–like" particles.

[2] Particles seen in the laboratory can also be bound states of fundamental fields.

served for fundamental particles indicates the occurrence of Electro-Weak Symmetry Breaking (EWSB). After EWSB, masses are given to the W and Z bosons through the Higgs mechanism, and to other particles via Yukawa couplings. The existence of fundamental Higgs bosons is the simplest explanation for EWSB.

A Standard Model Higgs boson field has properties that immediately hints of physics beyond the Standard Model. One possibility is *Supersymmetry*, which postulates a symmetry of space–time between integer and half–integer spin particles. This solves many problems with particle theory at high energy, such as the stability of the Higgs boson mass scale to radiative corrections and the unification of the gauge couplings. Also, Supersymmetry is required for Superstring theory, the leading theory for unifying gravity with the other forces.

The Higgs sector in the context of Supersymmetry is more complicated than in the SM, since (at least) two doublets of fields are required, instead of just one. $\tan \beta$ is the ratio between the vacuum expectation values (VEV) of the two doublets. Five Higgs bosons remain after symmetry breaking, three of which are neutral. This thesis presents a search for the neutral Higgs bosons in the Supersymmetric Standard Model. The data for the search was collected using the DØ detector, which records proton on anti–proton ($p\bar{p}$) collisions at a center–of–mass energy of 1.96 TeV from the Tevatron particle accelerator at Fermilab National Accelerator Laboratory, outside Chicago, IL. Theory predicts that neutral Higgs bosons will be produced in these collisions and their decay products observable if their masses are small enough (<150 GeV/c^2) and $\tan \beta$ is large (>50). Such a discovery would validate a model which is thought to give all known particles their mass, and point the way towards new physics which would become important at higher energies.

Currently, the most restrictive limits on properties of Higgs bosons are from the LEP II experiments at CERN. These experiments have excluded a CP-odd neutral Higgs boson with mass <91 GeV.

The research completed for this thesis is the first search for neutral Higgs bosons in

the context of Supersymmetry at the Tevatron in Run II, which began in March 2001. In addition, the work expands on previous methods. A dedicated multi–jet trigger is optimized for signal efficiency, which adds to the overall sensitivity significantly. The simulated Higgs boson signals are compared in detail to modern theoretical calculations. The heavy–flavor multi–jet backgrounds are compared to simulations of the SM for the first time, and novel methods for normalizing these backgrounds to data are applied. Lastly, limit–setting methods are used which take advantage of the full shape of the simulated signals and backgrounds, including the natural width of the Higgs bosons at high $\tan\beta$.

Chapter 2 describes the theory of the Standard Model, and Higgs bosons (particularly in the context of Supersymmetry) in greater detail. Chapter 3 documents the apparatus used for the experiment, the DØ detector and the Fermilab accelerators. Chapter 4 discusses the techniques used to reconstruct the collected data and to calibrate the detector's responses. Chapter 5 details the methods used to simulate the signals expected from neutral Higgs bosons as observed by the detector, as well as the simulated backgrounds which are used to understand the data that is observed. Chapter 6 explains the trigger used to collect the events of interest to this analysis and presents the resulting data sample. Chapter 7 describes the data analysis methods used to search for a Higgs boson signal. Finally, Chapter 8 reports and discusses the results obtained from this search.

4

Chapter 2

THEORY

2.1 The Standard Model

All of the known fundamental [1] particles and their interactions (except gravity) are described by a single theory, the Standard Model (SM). The SM is a Quantum Field Theory (QFT), which incorporates the two great physical theories of the 20th century: Relativity and Quantum Mechanics [1]. The theory is therefore valid at all relative velocities, up to the speed of light, and also at small scales, much smaller than the size of a single proton. Gravity is not included in the SM, but is instead described by a Classical Field Theory, General Relativity. The SM does not describe reality at extremely small (or fast) scales or at extremely high energy, where the effects of gravity become large. However, the energy at which gravity becomes important is predicted to be far greater than the energy probed at our current accelerators, by many orders of magnitude [2].

2.1.1 The Three Families of Matter

A beautiful structure organizes the fundamental particles. They are grouped into three *families*, each family being identical except for the masses of its members. (The

---

[1] A particle is called "fundamental" if there is no known substructure to it, i.e. it can not be understood as being a bound state of smaller things. Practically speaking, a fundamental particle means that its field is written in the Lagrangian. This latter definition is more in line with the idea that the Standard Model is a low–energy "effective" field theory.

[2] Theories of Large Extra Dimensions (LED) bring the scale of where gravity becomes important down to the energies we can probe with accelerators, but this is a separate topic of search for new physics.

Table 2.1: The "matter" particles of the Standard Model, and their interactions: strong force (S), weak force (W), and electro–magnetic force (E). (All particles interact via gravity.) Each family contains an up–type quark, down–type quark, charged lepton, and a neutrino.

Family 1	Family 2	Family 3	Interactions
up quark (u)	charm quark (c)	top quark (t)	S W E
down quark (d)	strange quark (s)	bottom quark (b)	S W E
electron (e)	muon (μ)	tau (τ)	W E
electron neutrino (ν_e)	muon neutrino (ν_μ)	tau neutrino (ν_τ)	W

different masses are presumed to arise from interactions with the Higgs boson, as will be described below.) Each family contains two quarks (an "up" and "down" type), a charged lepton (like the electron), and a neutrino. All members of a family have spin 1/2, and are thus *fermions*, meaning they have half–integer spin and obey Fermi–Dirac statistics. Force–carrying particles, described below, have spin 1, and are thus *bosons*, meaning that they have integer spin and obey Bose–Einstein statistics.

The three families of fermions in the SM are shown in Table 2.1. The first family contains the particles which make up nearly all the visible matter around us, i.e. atoms. The up and down quarks form protons and neutrons, which bind to form the nuclei of all atoms. The electrons create the clouds of charge around the nuclei. Because they are fermions, only one electron can occupy a given quantum state at a time, thus intricate patterns are formed which give rise to chemistry and the complicated spectra of light emitted from atoms. The electron neutrino does not help to form atoms. It is released in the decay of neutrons and participates in reactions which are crucial for powering the Sun.

The other two families contain particles which are heavier. They are able to decay, through interactions of the SM, into particles of the first family. The second family contains the strange and charm quarks, the muon, and the muon neutrino. The third

family is heavier still, and contains the top and bottom quarks, the tau, and the tau neutrino. While these particles do not naturally occur in the world around us (except in "cosmic rays"), they are produced in collisions at the Tevatron.

2.1.2 The Three Forces

All of the forces, or interactions, between these fundamental particles (except gravity) are described by three simple *local gauge symmetries*. A local gauge symmetry is an internal symmetry with respect to some group of parameters which is preserved at all points in space (and time) by fields of a QFT. Changing (or *rotating*) the values of the parameters independently at each point leaves the results of any physical measurements unchanged.

For instance, the Electro-Magnetic (EM) force is responsible for the attraction of oppositely charged particles, magnetic fields, and light (EM radiation). It is carried by the photon and results from a $U(1)$ local gauge symmetry of the complex phase of the fields. Demanding that no observable change results from changing this phase at each point in space–time independently creates what we observe as Electro-Magnetism. The requirement of the $SU(2)$ and $SU(3)$ local gauge symmetries creates what we observe as the Weak and Strong forces, respectively. The Weak force is responsible for nuclear β–decay and is transmitted by the W and Z bosons. The Strong force is responsible for holding protons, neutrons, and other strongly–interacting particles together and is exchanged via bosons called gluons.

2.1.3 The Higgs Mechanism

As described above, fermions are observed to have mass. In fact, the masses of the fermions allow the three families to be distinguished from each other [3]. Also, the W and Z bosons are massive, which explains why the force carried by them is "weak":

---

[3]The only parameters of the SM dependent on family are the couplings of the fermions to the Higgs fields.
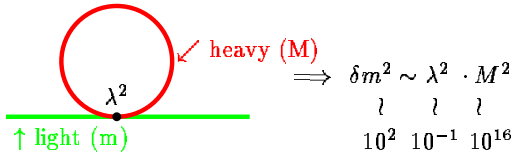
Figure 2.1: Radiative corrections to the Higgs boson mass by Standard Model fields. A Higgs boson with small mass (m) and coupling λ to a heavy fermion with mass (M) receives a large contribution to its mass, δm.

fermions must come very near to each other in order to interact via the Weak force. The bosons transmitting the force are heavy and short–lived, thus the force is short–ranged. The massless photon has an infinite lifetime and thus and communicates the EM force over an infinite range.

Masses arise in the SM through the Higgs field, which is a scalar (spin–0) complex–doublet quantum field. The Higgs field is very special, since it takes on a non–zero Vacuum Expectation Value (VEV). In its lowest energy state (the vacuum), the Higgs field has positive energy density [4]. This breaks the Electro–Weak symmetry which would exist if all fields had zero energy in the vacuum. The W and Z bosons each absorb a component of the Higgs field as their longitudinal component, giving them masses [5]. The fermions acquire masses through Yukawa interactions with the Higgs field. Through the same interactions which give rise to the masses of the fermions and W/Z bosons, Higgs particles (excitations of the Higgs field above its vacuum state) can be created.

2.2 Supersymmetry

The SM is a consistent theory of fundamental particles and their interactions up to the energies they have been studied. However, the SM is immediately seen to have some short–comings. The Higgs field is a scalar, and thus can undergo radiative quantum corrections, from any heavy particles of the type shown in Figure 2.1 (from [8]). In principle, these corrections diverge to infinity, since they are quadratic in nature and the momentum of the particle in the loop is unconstrained. Even if the momentum is limited to some very high energy, to represent an ignorance of very high–energy physics, the corrections are still very large. The radiative corrections would naturally drive the Higgs mass to the scale where gravity becomes important (about 10^{19} GeV). This is known as the *hierarchy problem*: why the VEV of the Higgs field is so low compared to where it would naturally be driven by radiative corrections.

Supersymmetry provides a partial solution to the hierarchy problem. A new bosonic partner to each fermion of the SM is introduced, with degrees of freedom which precisely cancel the radiative corrections to the Higgs VEV from the fermion fields (see Figure 2.2, from [8]). Supersymmetry requires the masses of the partners to be the same as those for the SM particles. Since these Supersymmetric partners have not yet been observed, Supersymmetry must be broken, such that the partners are heavier than the SM particles. If the differences in mass between SM particles and their partners is too large, then the hierarchy problem returns, since the partners would be too heavy to cancel the loops from the SM particles. Thus, the hierarchy problem has really been re–phrased, asking why Supersymmetric partners are just slightly larger than the SM particles. Progress has been made, however, because the re–phrased question has hope of being answered, whereas the original hierarchy problem by definition can not be.

---

[4]The Higgs field gives no directional preference to space because it is scalar.

[5]Massless gauge bosons (such as the photon) have transverse components only.
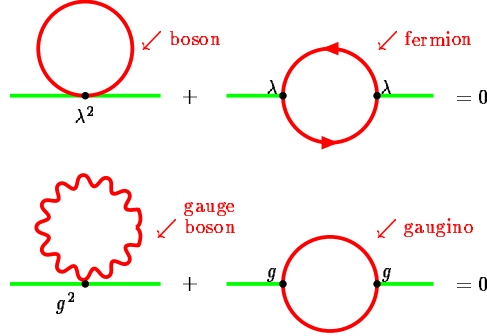
Figure 2.2: Cancelation of radiative corrections to the Higgs boson mass by Supersymmetric degrees of freedom. For each bosonic degree of freedom in the SM, there exists a fermionic degree of freedom from Supersymmetry, with exactly the coupling necessary, λ, to cancel the radiative corrections. Similarly, for each gauge boson degree of freedom in the SM, there exists a gaugino degree of freedom from Supersymmetry, with exactly the coupling necessary, g, to cancel those radiative corrections.

Supersymmetry has other attractive features as well. Unification of the three gauge symmetries into a larger, single gauge group (such as $SU(5)$), could provide explanations for many coincidences of high–energy theory. However, this requires the coupling constants of the three forces to be identical at some high energy. The coupling constants vary as a function of energy through renormalization and screening effects. In the SM, the forces do not unify at any energy, but for the field content of Supersymmetric extensions of the SM, they do (see Figure 2.3). In addition, the energy at which the couplings unify is at an energy large enough to predict the rate of proton decay to be below experimental limits. Supersymmetry is also needed for unifying gravity with the other forces. It is the only consistent mathematical framework [6] which can accommodate both spin–2 particles (such as the graviton) and spin–1 gauge bosons (such as the photon). Supersymmetry also unifies the forces (spin–1 and spin–2 fields) with matter (fermions) and Higgs fields (scalar bosons).

---

[6] a unique Lie algebra

Figure 2.3: Unification of the coupling constants for the three gauge symmetries as a function of energy in the SM (left) and MSSM (right).

2.3 The Higgs Bosons in Supersymmetry

Supersymmetry is more restrictive than non–Supersymmetric extensions of the SM, and does not allow a single[7] Higgs field. An even number of Higgs fields must be present for Supersymmetric theories to be consistent [8]. The simplest scenario is thus *two* Higgs fields, which happens also to give the best unification of the couplings at high energy. There are thus *five* Higgs bosons remaining after EWSB, 2 neutral CP–even scalars, h and H (where H is defined to be the heavier state), a neutral CP–odd scalar, A, and two charged states, $H^{\pm}$. The ratio of the vacuum expectation values of the two Higgs fields is called $\tan\beta$.

Also, Supersymmetry imposes relations between the components of the fields,

---

[7]A "single" Higgs field should be taken to mean a complex doublet field, having 4 degrees of freedom.

[8]An even number of Higgs doublets is needed to cancel quantum anomalies, for instance. At least two doublets are also required to give masses to both the up– and down–type quarks

such that only two free parameters remain in the Higgs sector. The typical input parameters chosen are m_A, the mass of the CP–odd Higgs boson, and $\tan\beta$. Given these, the masses of the other Higgs bosons and all couplings to fermions can be derived [5]. In general, the coupling of the neutral Higgs bosons to the down–type quarks, such as the b–quark, are roughly proportional to $\tan\beta$, and thus production cross sections are proportional to $\tan^2\beta$. The Higgs bosons' widths are relatively small (compared to the resolution possible with modern particle detectors) up to very high $\tan\beta(\lesssim100)$. The neutral Higgs bosons decay about 90% of the time to a pair of bottom quarks, with decays to τ leptons making up almost all of the remainder.

At leading–order (LO), the following relationships may be derived: $m_h < m_Z|cos(2\beta)|$, $m_h < m_A$, $m_H > m_A$, and $m_{H^\pm}^2 = m_W^2 + m_A^2$. However, large radiative corrections from virtual top quark, stop quark (the Supersymmetric partner of the top quark), and bottom quark (at high $\tan\beta$) loops extend the upper limit of m_h from m_Z (91 GeV/c^2) to about 135 GeV/c^2. The dependence of m_h and m_H on m_A is shown on the left sides of Figures 2.7, 2.8, 2.9, and 2.10 for several values of $\tan\beta$. At high $\tan\beta(\gtrsim20)$, the A is always nearly degenerate in mass with either the h or the H.

The coupling of the neutral Higgs bosons to up–type quarks, down–type quarks, leptons, and vector bosons is strongly dependent on $\tan\beta$. The coupling of the h to the charged leptons and down–type quarks is a factor of $-sin\alpha/cos\beta$ larger than its SM value, the coupling to the H is increased by a factor of $cos\alpha/cos\beta$, and the coupling to A is directly proportional to $\tan\beta$. The parameter α is a quantity derived from m_A and $\tan\beta$ which describes the mixing between the h and the H:

$$\cos^2(\beta - \alpha) = \frac{m_h^2(M_Z^2 - m_h^2)}{m_A^2(m_H^2 - m_h^2)} \tag{2.1}$$

While $\tan\beta$ is a free parameter of the model, there is good reason to believe that it is large. A $\tan\beta$ value of ~35 would explain naturally the ratio of the top to bottom quark mass. Both the top and bottom quarks would have couplings very near unity
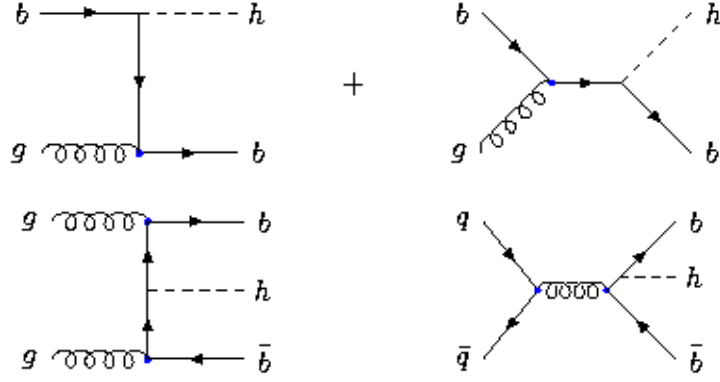
Figure 2.4: Leading–Order Feynman diagrams showing gb→bh production (top) and gg,$q\bar{q}$→$b\bar{b}$h production (bottom).

to their Higgs fields. But the down–type Higgs field would have a VEV lower by a factor of $\tan\beta$, endowing the down–type quarks with proportionally smaller masses. This solution would simplify the inclusion of up– and down–type quarks into a Grand Unified Theory. Also, Supersymmetry can provide a candidate particle for the dark matter in the Universe. The interactions of neutralino dark matter particles, and thus their annihilation rates in the early Universe, are strongly dependent on $\tan\beta$. A high value of $\tan\beta$ most naturally provides the best agreement with the cosmologically observed dark matter content of the Universe [7].

2.3.1 Production

This search looks for neutral Higgs bosons produced in association with bottom quarks. Associated production of neutral Higgs bosons with either one or two high p_T (>15GeV/c) b–quarks takes place through the LO processes gb→bh and gg,$q\bar{q}$→$b\bar{b}$h, respectively, as shown by Feynman diagrams in Figure 2.4. Identical diagrams also exist for the H and A.

Figure 2.5 shows the cross–sections for A and h/H production associated with a $b\bar{b}$ pair with $\tan\beta$= 1.5 and 30 at a $p\bar{p}$ collider at $\sqrt{s} = 1960$ GeV, calculated with

the Hqq program [9]. Figure 2.6 shows the same at $\tan\beta = 5$ and 60. At high $\tan\beta$, the production of either the h or the H is always nearly equal to that of the A, for all m_A. Since this analysis is unable to distinguish between the h/H and the A, we will simply assume that production of the A doubles the total cross–section from that of the h or H alone. At no point in parameter space is production of all three neutral Higgs bosons enhanced compared to SM Higgs production.

2.3.2   Masses, Widths, and Branching Fractions

All widths, branchings, and masses for Higgs bosons have been calculated using the program HDECAY [10]. The widths of the CP even and odd neutral Higgs bosons for $\tan\beta$ of 30 are shown in Figure 2.9 (right side). They are smaller than the detector resolution ($\sim$20 GeV) for $\tan\beta \lesssim 100$. When the production cross–section of either the h or H is very small, its width is also very small. Both effects are caused by the decreased coupling of the boson to the bottom quark. The widths of the neutral Higgs bosons are directly proportional to their coupling to the bottom quark, which is in turn proportional to $\tan\beta$. More plots of cross section and widths can be found below in Figures 2.7, 2.8, 2.10, and 2.11 for $\tan\beta$ of 1.5, 5, 60, and 120.

The branching fractions of the neutral Higgs bosons into $b\bar{b}$, τ's, and other channels are shown in Figure 2.14 for $\tan\beta$ of 30. (Branching fractions are also shown in Figures 2.12, 2.13, 2.15, and 2.16 for $\tan\beta$ of 1.5, 5, 60, and 120.) The neutral Higgs bosons all decay to $b\bar{b}$ $\sim$90% of the time at high $\tan\beta$ whenever their production in association with bottom quarks is enhanced, because the coupling of the Higgs to the bottom quark is enhanced. The τ channel is also very interesting and will be pursued in a future analysis.

14
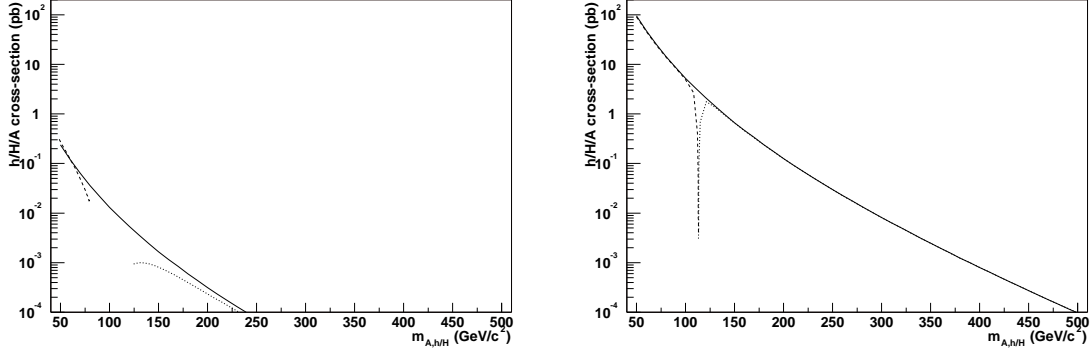
Figure 2.5: Production cross–sections for neutral Higgs bosons associated with a $b\bar{b}$ pair at Leading–Order, for $\tan\beta$ of 1.5 (left) and 30 (right). (The A is solid, the h is dashed, and the H is dotted.)
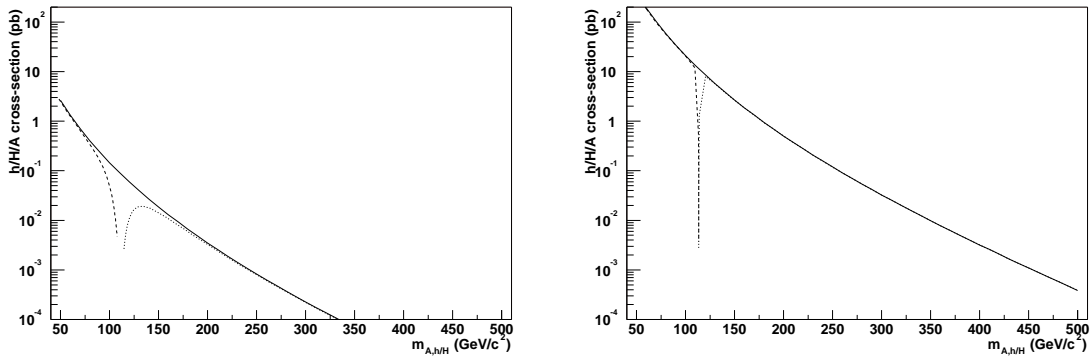
Figure 2.6: Production cross–sections for neutral Higgs bosons associated with a $b\bar{b}$ pair at Leading–Order, for $\tan\beta$ of 5 (left) and 60 (right). (The A is solid, the h is dashed, and the H is dotted.)
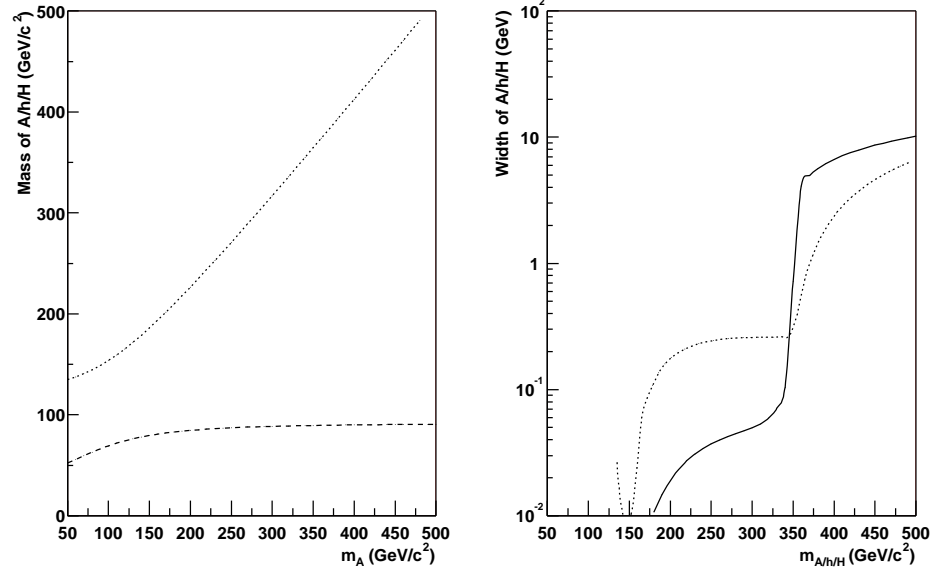
Figure 2.7: Dependence on m_A of m_h and m_H (left). Total widths for neutral Higgs bosons (right). Both are for $\tan\beta$ of 1.5.
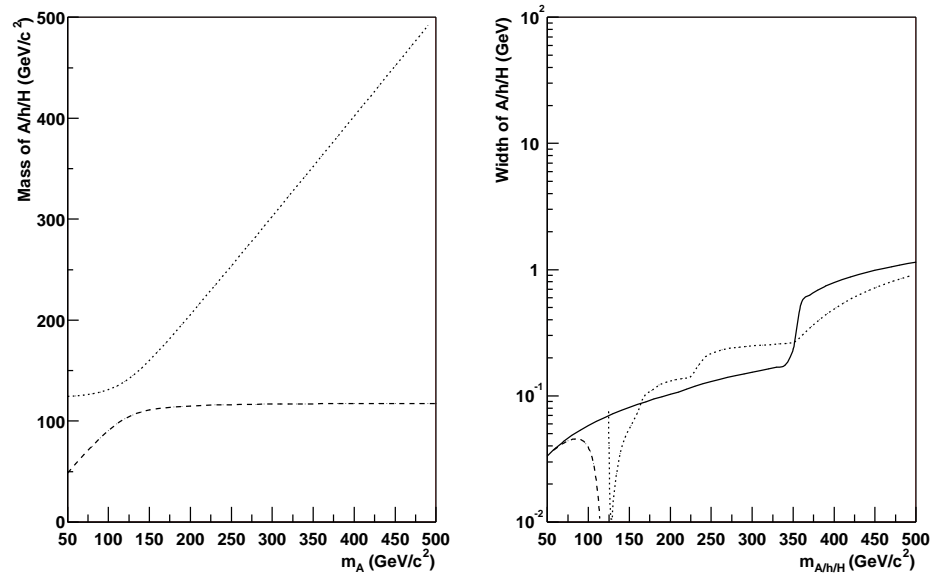
Figure 2.8: Dependence on m_A of m_h and m_H (left). Total widths for neutral Higgs bosons (right). Both are for $\tan\beta$ of 5.
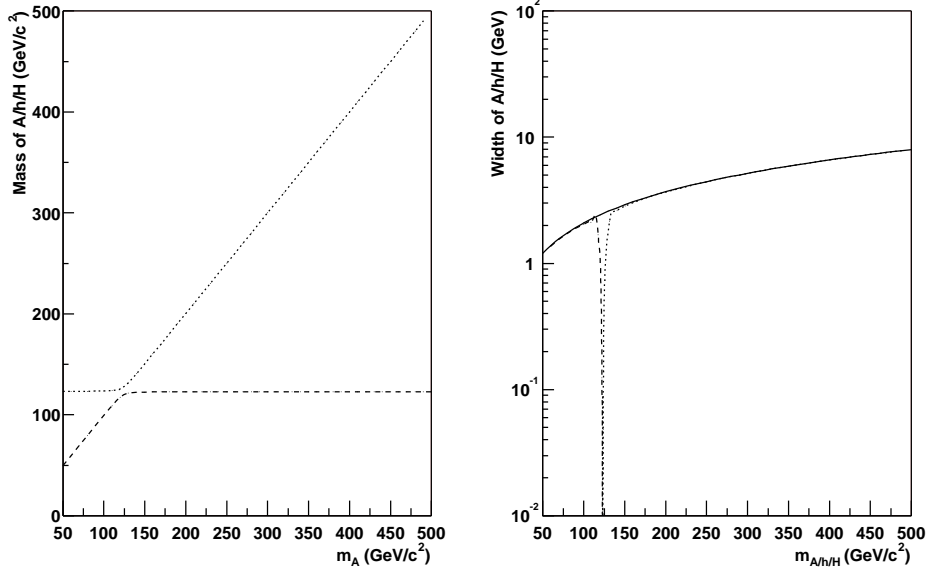
Figure 2.9: Dependence on m_A of m_h and m_H (left). Total widths for neutral Higgs bosons (right). Both are for $\tan\beta$ of 30.
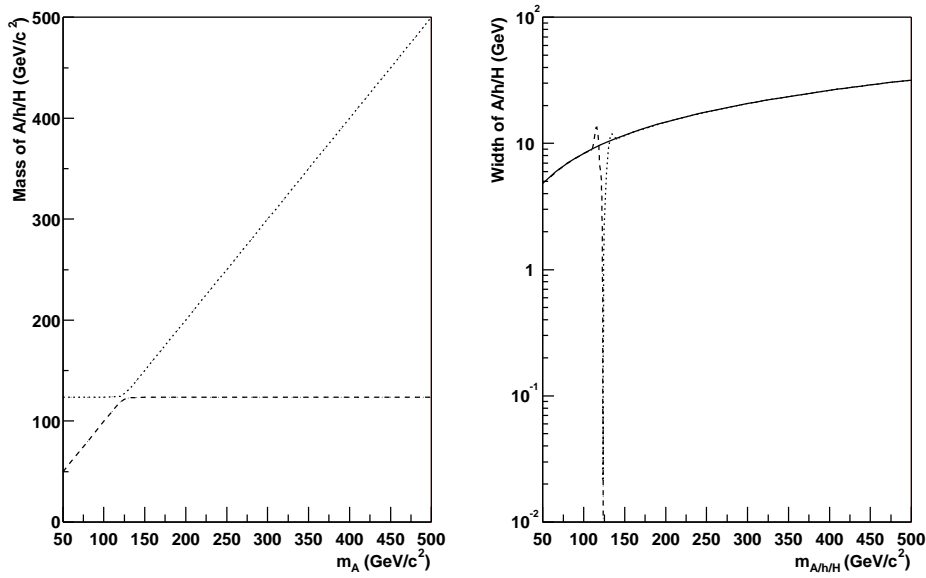
Figure 2.10: Dependence on m_A of m_h and m_H (left). Total widths for neutral Higgs bosons (right). Both are for $\tan\beta$ of 60.
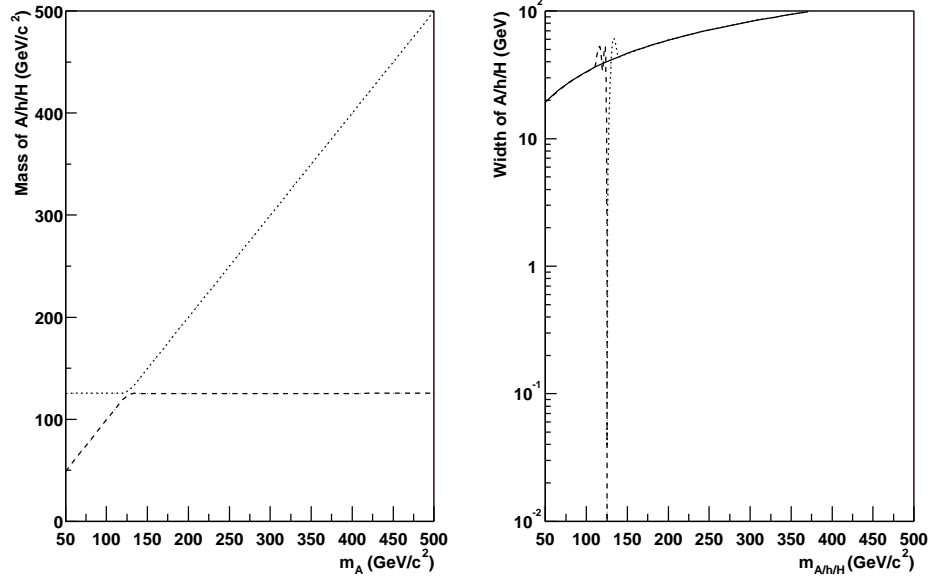
Figure 2.11: Dependence on m_A of m_h and m_H (left). Total widths for neutral Higgs bosons (right). Both are for $\tan\beta$ of 120.
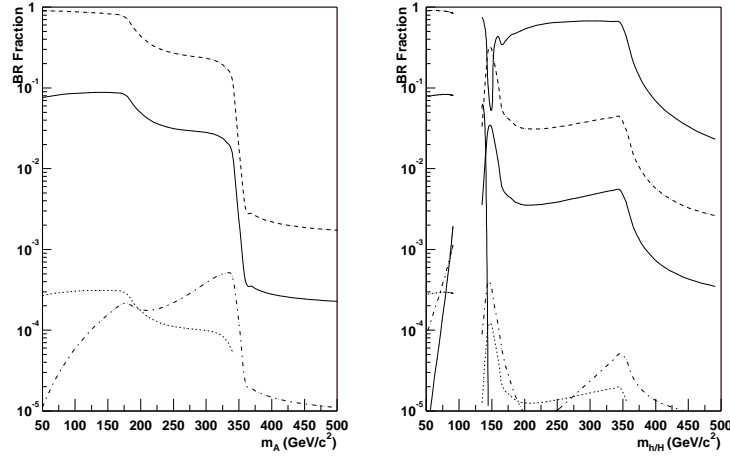
Figure 2.12: Branching fractions for neutral Higgs bosons (A on the left, h/H on the right) at $\tan\beta$=1.5. $b\bar{b}$ is dashed (highest), τ is solid (middle), and μ is dotted (lowest). Also shown is the decay of the light Higgs to WW, which turns on at very high m_A, and the decay of H to hh which turns on at twice m_h.
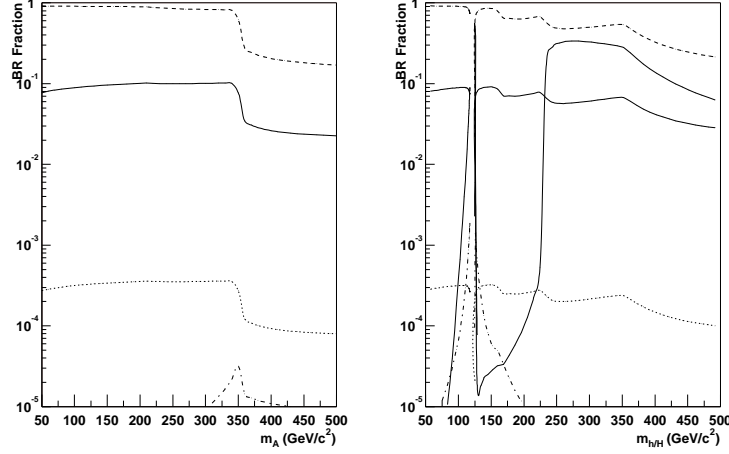
Figure 2.13: Branching fractions for neutral Higgs bosons (A on the left, h/H on the right) at $\tan\beta$=5. $b\bar{b}$ is dashed (highest), τ is solid (middle), and μ is dotted (lowest). Also shown is the decay of the light Higgs to WW, which turns on at very high m_A, and the decay of H to hh which turns on at twice m_h.
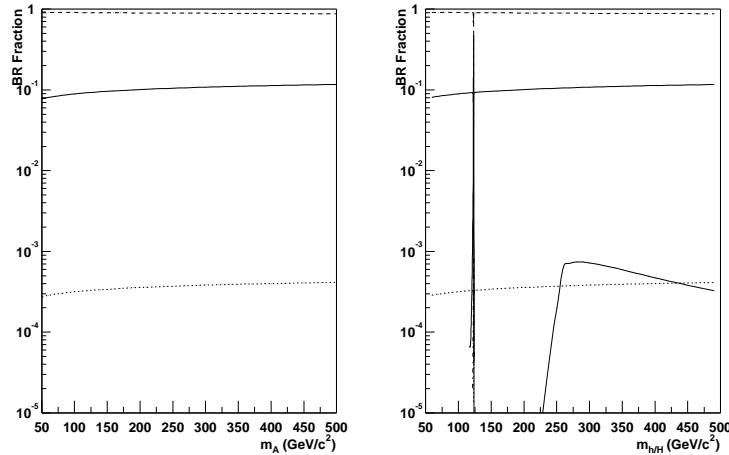
Figure 2.14: Branching fractions for neutral Higgs bosons (A on the left, h/H on the right) at $\tan\beta$=30. $b\bar{b}$ is dashed (highest), τ is solid (middle), and μ is dotted (lowest). Also shown is the decay of the light Higgs to WW, which turns on at very high m_A, and the decay of H to hh which turns on at twice m_h.
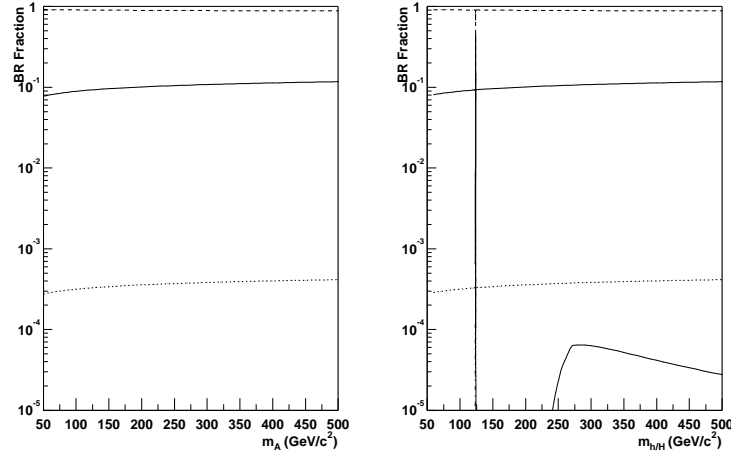
Figure 2.15: Branching fractions for neutral Higgs bosons (A on the left, h/H on the right) at $\tan\beta$=60. $b\bar{b}$ is dashed (highest), τ is solid (middle), and μ is dotted (lowest). Also shown is the decay of the light Higgs to WW, which turns on at very high m_A, and the decay of H to hh which turns on at twice m_h.
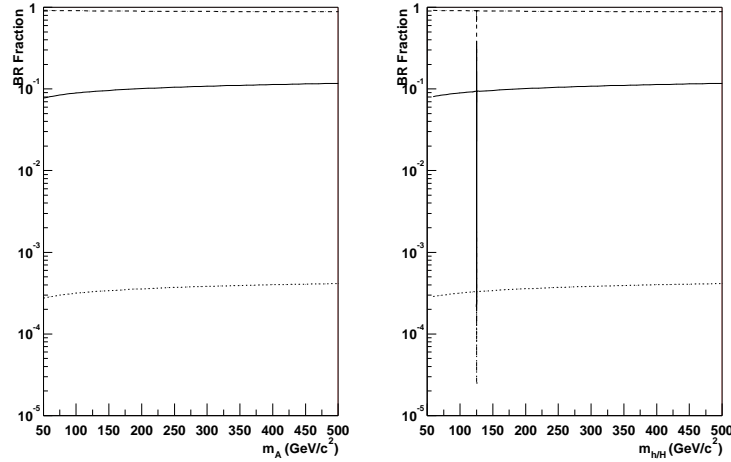
Figure 2.16: Branching fractions for neutral Higgs bosons (A on the left, h/H on the right) at $\tan\beta$=120. $b\bar{b}$ is dashed (highest), τ is solid (middle), and μ is dotted (lowest). Also shown is the decay of the light Higgs to WW, which turns on at very high m_A, and the decay of H to hh which turns on at twice m_h.

Chapter 3

EXPERIMENTAL APPARATUS

The data for this research was recorded with the DØ detector during Run II of the Tevatron at Fermilab. Run I took place during 1992-1995. The goal of Run II, which began in 2001, is to deliver about 100 times the number of collisions to the experiments by 2009, and at a slightly higher energy (1.96 TeV as opposed to 1.8 TeV) [1]. In addition, the detectors have been significantly upgraded for Run II to enhance their capability to observe interesting physics. At DØ, the major upgrades used for this analysis are the new triggering systems, and the addition of a central solenoidal magnetic field containing new tracking chambers including a high–resolution *vertex* detector.

This Chapter describes the basics of the operation of the Fermilab accelerators used for Run II, $p\bar{p}$ collisions and how particles are observed in the laboratory, and details of the pieces of the DØ detector which are relevant to this analysis.

3.1 The Fermilab Accelerators

The largest Fermilab accelerator, the Tevatron, is a 1 km radius synchrotron made out of about 1000 superconducting magnets, which are able to accelerate and hold a 980 GeV beam of protons and anti–protons (circulating in the opposite direction) and collide them head on. The collisions occur at two beam crossing interaction regions, one of which is at the center of the DØ detector. The interaction regions have a 3D–Gaussian shape and a width of about 30 cm along the beam axis (the "z" direction),

---

[1]The increase in center of mass energy for Run II, though modest, increases the production rates of heavy particles, such as pairs of top quarks, by about 40%.

and about 30 μm in the transverse directions. Bunches of protons and anti–protons cross almost every 396 ns [2].

The Tevatron does not simply accelerate protons and anti–protons from rest. Instead, many stages of acceleration and storage prepare the protons and anti–protons for injection into the Tevatron. The various components are diagrammed in Figure 3.1. Protons are first accelerated as H$^-$ ions using a Cockcroft-Walton device to 0.75 MeV, about 30 times the kinetic–energy of electrons inside an old–fashioned Cathode–Ray Tube (CRT) television. A linear accelerator about 500 feet long consisting of RF cavities is then used to accelerate the ions to 400 MeV, after which they pass through a carbon foil, which strips off the electrons leaving only the protons. These protons then enter the Booster, a circular synchrotron about 500 feet in diameter, where they are grouped into bunches and accelerated to 8 GeV. Proton bunches are injected into the Main Injector, where they are accelerated to 150 GeV. To make anti–protons, proton bunches from the Main Injector are focused onto a nickel target, and anti–protons are collected from the spray of particles created[3]. The bunch structure, remnant from the Main Injector proton bunches used to create the anti–protons, is removed in the Debuncher. The anti–protons are also stochastically cooled, their energies are made more uniform, in the Debuncher. The anti–protons are then transferred to the Accumulator for storage. When a sufficient number of anti–protons is present in the Accumulator, typically about 150–200 $\times 10^{10}$ anti–protons, 36 bunches of protons from the Main Injector are loaded into the Tevatron at 150 GeV. Then 4 bunches at a time of anti–protons are transferred to the Main Injector, where they are accelerated to 150 GeV and injected into the Tevatron. When 36 bunches of anti–protons have been injected, the Tevatron accelerates the protons and anti–protons to 980 GeV in one process. The beams are brought into focus in the collision regions, and the beam halos (protons and anti–protons in irregular orbits far from the beam

---

[2]Some bunch crossing intervals are intentionally left empty, for beam stability and calibration.

[3]About 15 anti–protons are collected from every million protons on target.
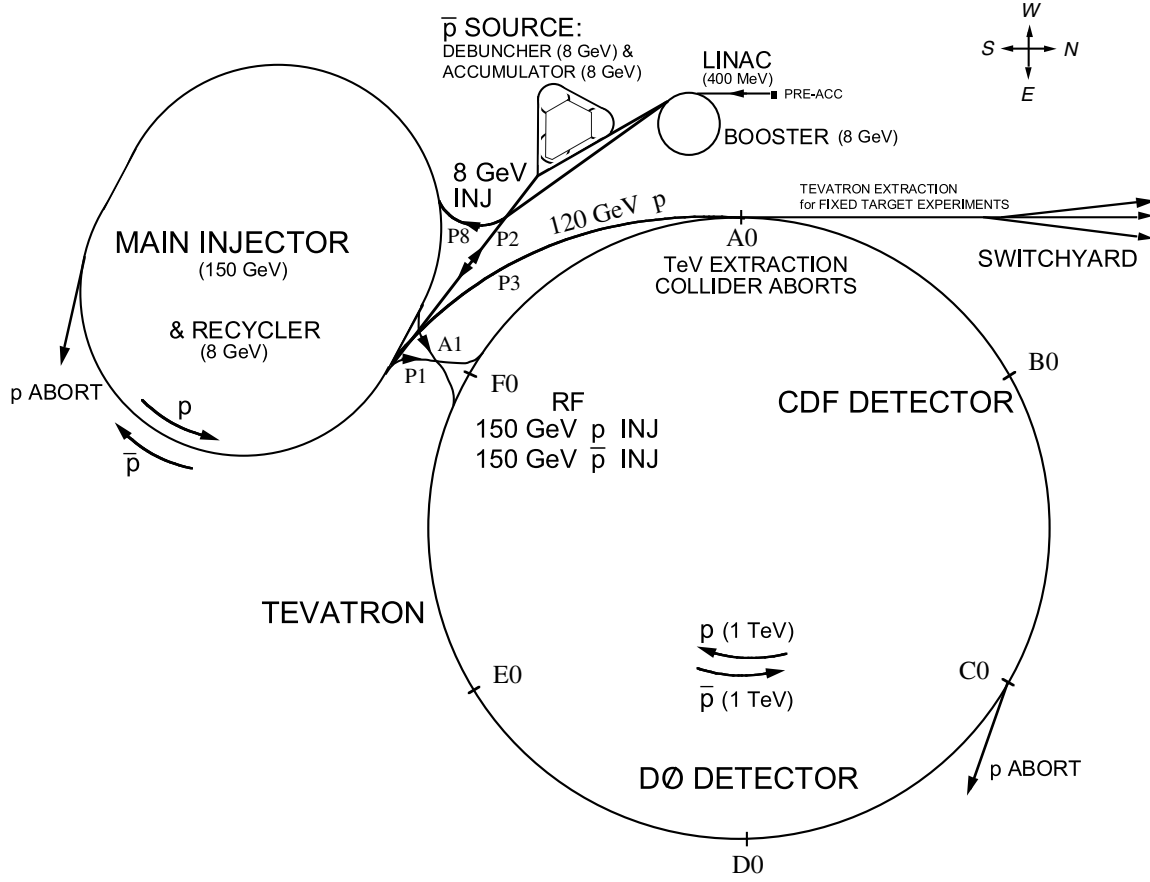
Figure 3.1: The Fermilab accelerator complex. Protons are accelerated in stages: starting in the LINAC, then to the Booster, the Main Injector, and finally into the Tevatron. Anti–protons are created using high–energy protons from the Main Injector focused on a fixed target, and they are collected and stored in the Debuncher and Accumulator. The Anti–protons are fed into the Main Injector and then the Tevatron, to collide head–on with protons at the CDF and DØ interaction regions.

center) are scraped away with collimators. At this point a *store* is declared and the collisions are recorded for typically about 20 hours. The beams are eventually lost due to a malfunction or dumped intentionally because of decreased beam currents and focus, which lead to exponentially decreasing luminosity (described below) as a function of time.

3.2 $p\overline{p}$ *Collisions and Particle Detection*

Most $p\overline{p}$ interactions at Tevatron energies simply scatter the particles at low angles. In the more interesting collisions, however, more energy is transferred between constituents of the two particles, and the proton and anti–proton are broken apart. A *parton* (a quark or gluon constituent) in the proton interacts directly with another parton in the anti–proton to create a *hard–scattering* reaction. The fragments of the broken proton and anti–proton continue nearly parallel to the beam-line, as small showers of color–neutral particles.

The hard–scattering reaction may result in intermediate resonances, such as a Higgs boson. The particles which leave the hard–scattering region can be any of those in the SM. However, only electrons, muons, neutrinos, photons, and a few kinds of strongly–interacting particles (hadrons) are semi-stable and live long enough to reach the detectors. The different semi-stable particles are measured in various ways, as described below.

Electrons and muons are charged, and leave energy in the tracking detectors. An externally applied solenoidal magnetic field bends the paths of the charged particles, allowing the particles' charge and momentum to be measured. Electrons and photons produce showers in the calorimeter, where their energy is measured. Hadrons produce showers deeper in the calorimeter, and their energy is measured. Muons escape through the calorimeter to the muon chambers where their momentum is measured using another (toroidal) magnetic field. Neutrinos escape the detector completely

and can only be partially reconstructed, through conservation of total (transverse) momentum.

Quarks (except top) and gluons which leave the hard–scattering region, do not live very long before they undergo *hadronization*, due to confinement, which is a non–perturbative aspect of the Strong force that does not allow color–charged particles to be isolated. The lower–energy state is for new particles to be created from the vacuum and be combined with the colored particles to create color–neutral bound states. This process creates a *jet* of particles, each traveling in the general direction of the initial quark or gluon. These jets are detected as broad showers of charged particles and energy deposited in the calorimeter.

3.3 Units and Coordinates

3.3.1 Luminosity

Modern particle experiments are often searching for rare processes, such as the production of Higgs bosons. The total number of times a given process occurs, N, is directly proportional to both the *cross–section* for the event type, σ, and the *integrated luminosity*, $\pounds$ [4].

$$N = \pounds \sigma \tag{3.1}$$

The cross–section is fixed for a given center–of–mass energy and particle beam type, parameters which are set by the accelerator design. The goal of the accelerator is thus to maximize the integrated luminosity delivered to the experiments. Particles that are typically collided have very small effective interacting areas, so the unit used for luminosity is also very small, the *barn*, which is 10^{-24} cm^2. Typical cross–sections for very interesting physical processes are usually of the order of pico–barns (pb), or 10^{-36} cm^2. Thus integrated luminosity is often measured in inverse pico–barns, pb^{-1}.

---

[4]The integrated luminosity is the integral with respect to time of the *instantaneous luminosity*, L. $\pounds = \int_t L$

3.3.2   Geometry

Standard spherical coordinates are used: r, ϕ (azimuthal), and θ. The system is centered at the interaction region, at the center of the detector. The z direction ($r = 0$) is defined as the beam–line. Instead of θ, the *pseudo–rapidity*, η, is often used, defined as:

$$\eta = -\ln(\tan\frac{\theta}{2}) \tag{3.2}$$

Pseudo–rapidity is the massless limit of *rapidity*, y:

$$y = \frac{1}{2}\ln(\frac{E + p_z}{E - p_z}) \tag{3.3}$$

which is invariant under boosts in the z direction. The number of particles which result from high–energy particle collisions is roughly constant as a function of η (assuming identical beam energies and particle types). Solid angles are often measured in terms of ΔR:

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{3.4}$$

which is approximately invariant under boosts in the z direction.

3.4   The DØ Detector

The DØ detector measures the properties of particles resulting from a high–energy $p\overline{p}$ collision at its center. A diagram of the DØ detector is shown in Figure 3.2. Multiple layers, of different detector types, measure various properties of the particles. At the center are the tracking chambers, which measure precisely the paths taken by the charged particles. The calorimeter surrounds the tracking chambers and records the energies of Electro-Magnetic and hadronic particles. Lastly, the muon chambers measure the momentum of charged particles which have escaped the calorimeter and passed through the thick iron toroid magnets. The transverse energy and direction of
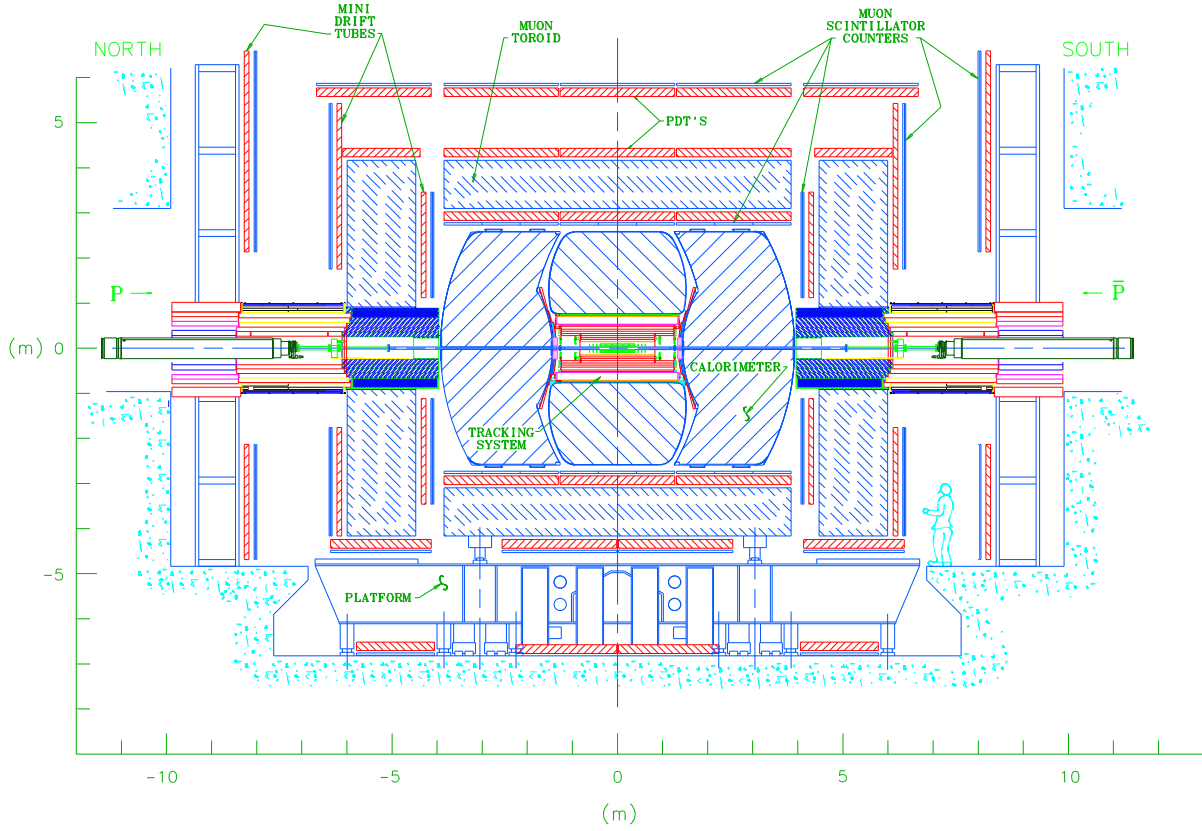
Figure 3.2: The DØ detector, upgraded for Run II. In this side view, proton bunches from the left collide with anti–proton bunches from the right in the center of the detector. The tracking chambers in the center surround the interaction region. The central and end–cap calorimeters are seen outside the tracking chambers. Outermost is the muon system which contains the thick iron toroid magnets.

neutrinos that were produced is reconstructed by demanding total transverse momentum conservation for each event[5]. Since collisions take place at rates far in excess of what can be recorded, an intricate, multi–layered, event trigger is employed to select only the most interesting events for complete reconstruction and analysis.

---

[5]The incoming protons and anti–protons have negligible momentum transverse to the beam, but the initial longitudinal momentum of the colliding partons varies greatly in each event.

3.4.1   Inner Detectors

The inner detectors are surrounded by a 2T superconducting solenoid magnet, which bends the paths of charged particles with a curvature inversely proportional to their momenta. Observing the curvature of a particle's path allows for a precise measurement of its momentum, as well as the sign of the particle's charge. A detailed diagram of the various components of the inner detectors is shown in Figure 3.3.

Luminosity System

The luminosity system is responsible for measuring the instantaneous luminosity being delivered to the experiment. The rate of inclusive inelastic $p\bar{p}$ scattering is measured by detecting charged particles from the interaction region. Since most inelastic $p\bar{p}$ interactions transfer a small amount of momentum between the proton and anti–proton, the particles resulting from the collision tend to be at large $|\eta|$ [6]. Two sets of plastic scintillator detectors are mounted on the inside face of the end–cap calorimeters (labeled "Level 0" on Figure 3.3). They are wedge–shaped and arranged symmetrically in ϕ around the beam pipe, from an $|\eta|$ of 2.7 to 4.4. The time resolution of the scintillator detectors is $\lesssim 0.2$ ns, needed to discriminate between particles originating from the interaction region from those in either of the beam halos remaining.

The Silicon Micro-strip Tracker

The detector nearest to the interaction region is the Silicon Micro-strip Tracker (SMT), which provides high resolution position measurements of the charged particle paths. These are used for determining whether any tracks came from *secondary vertices*, which are a good indication of the presence of bottom quarks, as will be discussed in Chapter 4.

---

[6]This is to be contrasted to high–energy $p\bar{p}$ collisions, which scatter outgoing particles more uniformly in η.
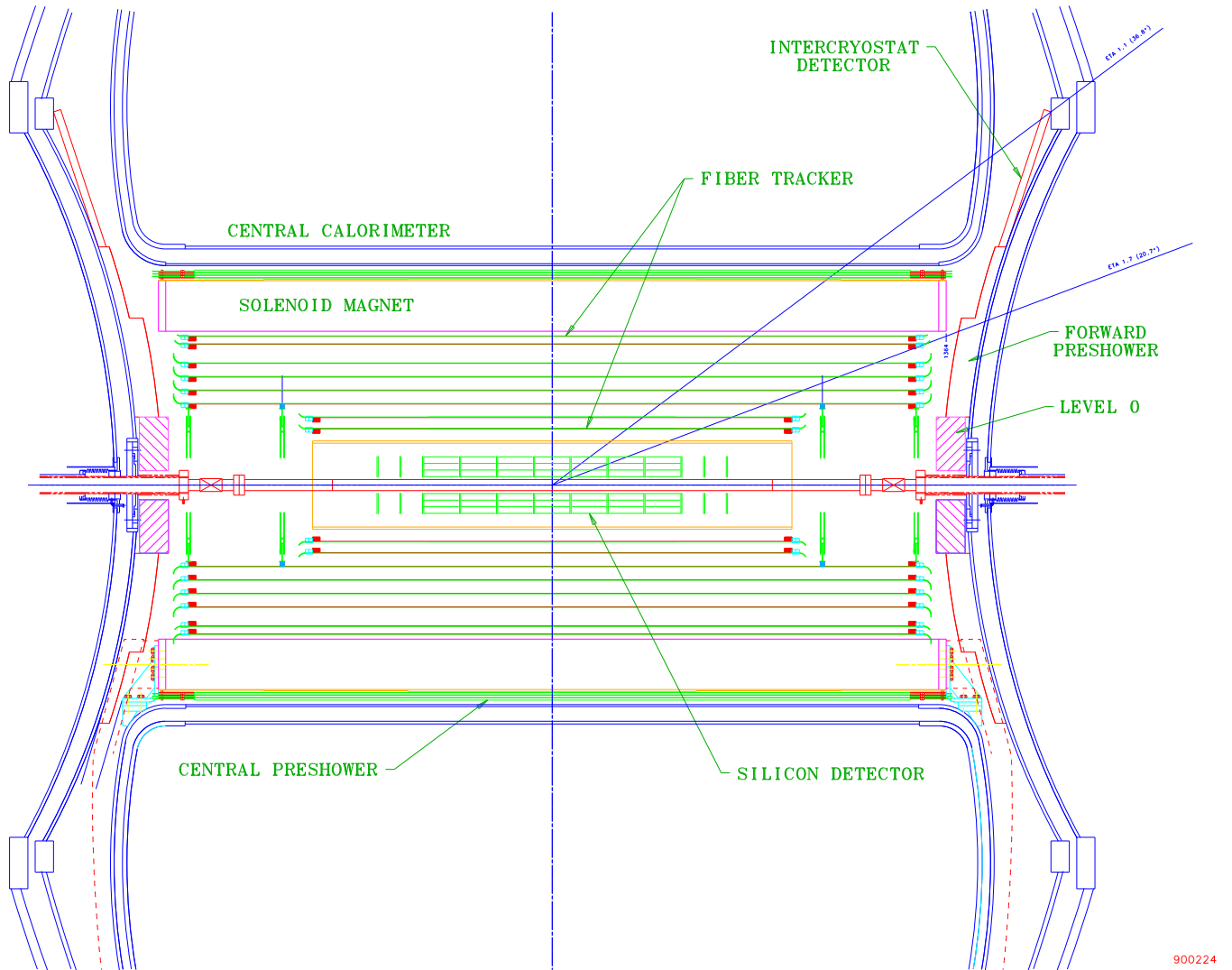
Figure 3.3: The inner detectors: the Luminosity System (Level 0), Silicon Detector (SMT), and Fiber Tracker (CFT).

A silicon detector makes use of the electron–hole pairs produced at a p–n junction in silicon when a charged particle passes through. These pairs are separated by an applied voltage and drift through the 300 μm thick wafers towards 50 μm wide conducting strips implanted in the silicon [7]. The charge collected from each strip is stored in a capacitor, until it is read–out and digitized by specialized electronics. Too much data is contained in the silicon strip capacitors to be digitized and read out on every event. Instead, an array of 32 capacitors per strip holds the analog charge until a Level 1 trigger accept (described in the Trigger System Section below) occurs. The charge for that event in the capacitor for each strip is then digitized, zero–suppressed [8], and transferred out of the detector.

The SMT detector geometry is shown in Figure 3.4. Six "barrels" surrounding the beam pipe make up the heart of the detector. Each barrel is 12 cm long and composed of 4 layers of silicon wafers and read–out chips, which are slightly overlapped, to prevent gaps in ϕ acceptance. Thus, a central ($|\eta|$<1.1) particle traverses at least 4 silicon detectors, with up to 8 possible. Due to inefficiencies, the typical number of reconstructed measurements is between 3 and 4 per central track.

Between the barrels (except the two most central) are placed "disks", perpendicular to the beam pipe, which runs through their centers. There are also 4 disks on each side of the 6 barrels. The disks are composed of 12 "F-wedges" each, which are silicon detectors that extend from a radius of 2.6–10.5 cm. There are also two larger "H-disks" placed at each end of the detector. They are made of 16 "H–wedges" each, extending from a radius of 9.5–26 cm. Together, these disks greatly extend the $|\eta|$ coverage of the silicon detector, out to about 3.0.

---

[7]The electrons and holes do not move directly towards the strips along the electric field lines, due to the presence of the 2T solenoidal magnetic field. The drift angles are corrected for during the SMT cluster reconstruction, discussed in Chapter 4.

[8]Zero–suppression is a simple data compression algorithm which only transmits the data from strips which are above threshold.
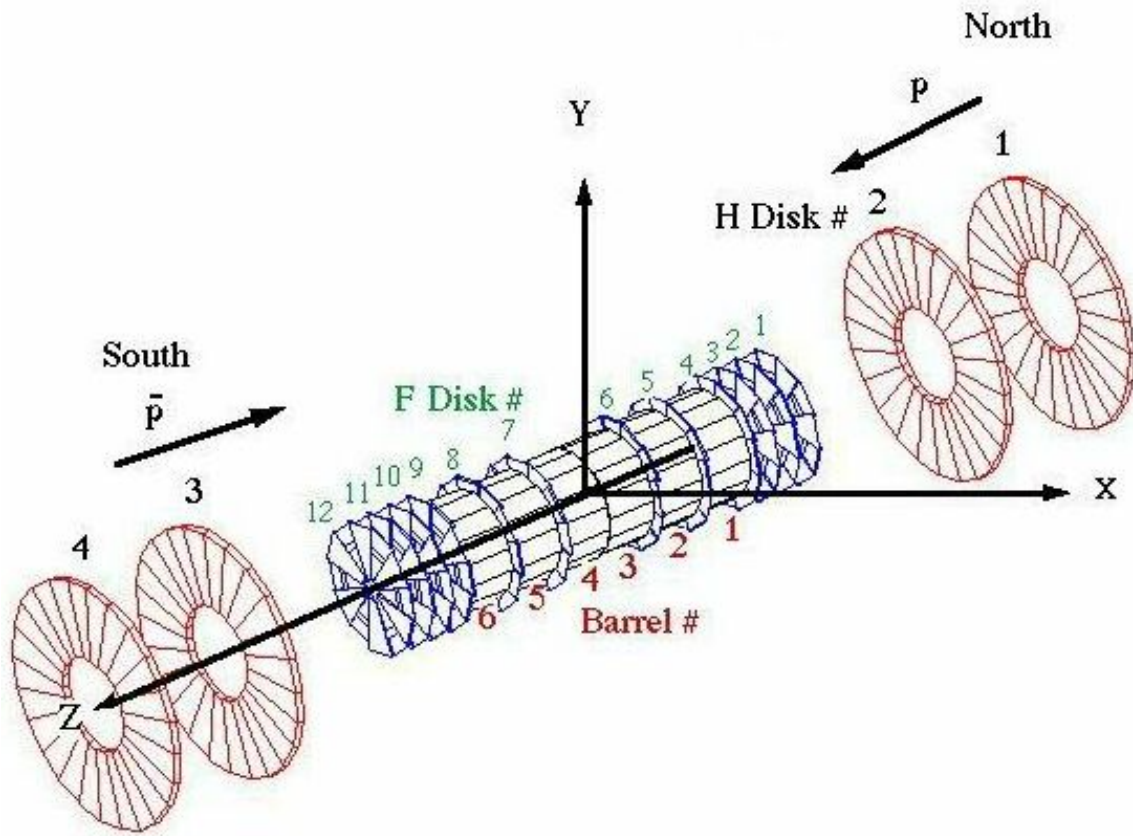
Figure 3.4: The Silicon Micro-strip Detector (SMT). Six barrels cover the central region. F–disks provide measurements for more forward tracks. H–disks extend the $|\eta|$ coverage out to about 3.0.

The Central Fiber Detector

Surrounding the silicon detector and extending out to the solenoid magnet, is the Central Fiber Tracker (CFT), a detector which aids the SMT in reconstruction of charged particle tracks. The CFT can detect charged particles up to $|\eta|$ of about 2. It makes use of 71,680 scintillating fibers, which contain a dye of molecules that are excited by high-energy charged particles. The molecules release photons in the visible part of the electromagnetic spectrum as they relax to their ground states, which are then detected.

Each scintillating fiber is 835 μm in diameter, composed of a 775 μm scintillating core surrounded by a coating with a high index of refraction, providing total internal reflection. The fibers are between 1.66–2.52 m long, and connect to clear wave-guides at one end, which carry the photons approximately 8–11 m to where they are detected. The other end of each fiber is coated with a surface which reflects photons back through the fiber into the clear wave–guide.

Each ionizing particle produces an average of about 10 photons in each fiber, which are detected using a Visible Light Photon Counter (VLPC) that converts the photons into an electrical pulse. The VLPC is a solid–state silicon device, 1 mm in diameter, operated at liquid Helium temperature to reduce electronic noise. Photons are converted into electron–hole pairs, and the holes create cascades of electrons due to the 6 V bias voltage applied. The electron cascade is detected as a current through the VLPC. A gain of over 50,000 is achieved and a quantum efficiency of about 80% per photon. The electrons from photon conversions are collected within $\lesssim$ 100 ns, before the next bunch crossing takes place. 1024 VLPC's are combined into "cassettes" which carefully regulate the temperature of the devices and contain the read–out electronics.

The CFT is constructed of 8 super–layers. Each super–layer is composed of two doublet–layers, an axial doublet–layer of fibers and a stereo doublet–layer which is at

a 3 degree angle relative to the beam axis. (The stereo doublet–layers alternate in the sign of the stereo angle.) A doublet–layer is composed of parallel, adjacent fibers bound into ribbons of 128 fibers each. A second layer is placed on top of the first, offset by half a fiber diameter, such that each fiber in the second ribbon maximally fills the space between the two fibers in the first ribbon. The efficiency of each doublet–layer is about 99% per particle, including dead channels. The digitization and read–out of the CFT is performed nearly identically to that of the SMT.

3.4.2 Calorimetry

The calorimeter is outside the solenoid magnet (and thus is in a region of low magnetic field). It measures the energy of photons, electrons, and hadronic jets of particles, by inducing them to create showers of energy using a large amount of dense material. The energy in the showers is sampled at many points, to determine its shape and energy. The central calorimeter extends out to $|\eta|$ of 1.1, and the forward calorimeters extend to $|\eta|$ of about 4.0. The regions between the central and forward calorimeters, called the Inter-Cryostat Region (ICR), is covered by special detectors, such that the coverage is nearly hermetic.

The showers are induced by layers of depleted Uranium [9]. Between the Uranium layers are 2.3 mm wide cells, which measure the ionization created by the showering particles in liquid Argon. A copper read–out pad in each cell is held at high voltage to create an anode which collects the ionization. The pad is insulated by a thick G10 coating, and an additional resistive coating. Thus the ionization creates a (reverse) image charge which builds up on the read–out pad.

The charge on each pad is sampled in analog form and fed into shaper and Base–Line Subtracter (BLS) boards which isolate the signal from the current beam–crossing

---

[9]Uranium is an ideal material because it is very nearly *compensating*, creating the same ionization per unit length from an incoming electron or pion (the most common hadronic particle in showers). Thus the ionization detected in a high–energy particle shower will be nearly the same, no matter how much of the shower energy goes into EM particles as opposed to hadronic particles.

from the ionization remaining from previous beam crossings. (The ionization takes a few μs to be completely absorbed, given the 430 ns drift time in Argon, as compared to 396 ns between beam crossings.)

A side view of the calorimeter geometry is shown in Figure 3.5. Radially, the calorimeters are composed of the electromagnetic layers (closest to the beam-pipe), the fine–hadronic layers, and the coarse–hadronic layer (outer-most). The electromagnetic–layers contain the electromagnetic showers, since their 65.6 mm total thickness of Uranium is over 20 electromagnetic interaction lengths. The 4 EM layers of cells are η x ϕ = 0.1 x 0.1 in size, except for the 3rd layer (from the beam-pipe), which has double the granularity (η x ϕ = 0.05 x 0.05) in the central region, to measure the shower shape better at its average shower maximum. The fine–hadronic calorimeter has 3 layers of cells in the central region and 4 in the forward. The cell size in this region is also η x ϕ = 0.1 x 0.1. The coarse–hadronic layer is the outer-most, and has half of the usual granularity (η x ϕ = 0.2 x 0.2). Together, the two hadronic calorimeters make up about 6.4 hadronic interaction lengths, thus containing nearly all of the hadronically showering particles.

The cells are arranged in a projective geometry in η, making it simple to measure the energy in a given η x ϕ region. All cells in each η x ϕ = 0.2 x 0.2 region are collectively called a *tower*. The total E_T in each tower is used for the calorimeter trigger (discussed below) as well as jet reconstruction.

3.4.3   Muon System

High–energy muons (with $p_T \gtrsim 3$ GeV/c) are measured in the muon system, which is the outermost layer of the detector (see Figure 3.6). A 2000 ton, 1.9 Tesla Iron toroid magnet bends the paths of the muons, and absorbs nearly all other particles. By measuring the muon path at three points, one before the toroid and two past it, the momentum of the muon can be ascertained with reasonable accuracy. The precise momentum of the muon is measured by matching the path of the muon to a high p_T
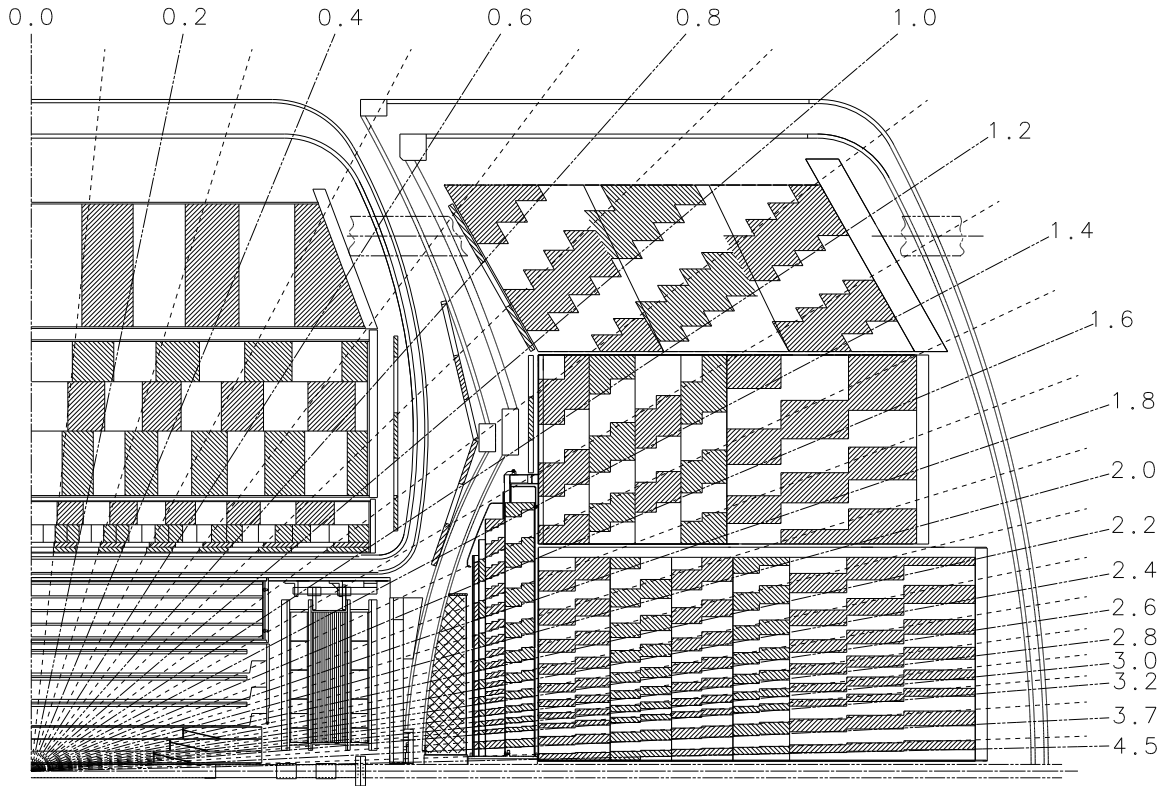
Figure 3.5: A side view of the calorimeters. The central calorimeter extends to $|\eta|$ of about 1.1, and the end–cap calorimeters cover $|\eta|$ up to about 4.0. The regions between the calorimeters, the Inter-Cryostat Region (ICR), are covered by special detectors, such that the coverage is nearly hermetic. Cells are arranged in a projective geometry of size $\eta \text{ x } \phi = 0.1 \text{ x } 0.1$.

track reconstructed in the inner tracking chamber. A muon also leaves about 3 GeV of energy in the calorimeter, distributed evenly along its path, which can also be used for muon identification.

The measurements in the muon system are made using drift–chambers, which extend out to $|\eta|$ of 2.0. The drift–chambers collect the ionization left by the muons in the organic – gaseous argon mixture onto gold wires held at high voltage. Position measurements are made by measuring arrival times of the ionization pulses, both relative to the beam crossing time, and at each end of the wire, allowing for a rough 3 dimensional position measurement. Additionally, scintillating plastic detectors on the outside of the muon detectors and on the inside of the first muon layers provide more spatial measurements of the muon path, and also gives very precise ($\lesssim$10 ns resolution) measurements of the arrival time of the particles, thus providing for good cosmic ray rejection.

3.4.4 Trigger Systems

As described above, bunches of protons cross with bunches of anti–protons nearly every 396 ns, or at a rate of about 2.5 MHz. However, only about 50 Hz can be recorded and analyzed, due to constraints of data storage and processing time required for reconstruction and analysis. Thus, only one out of each 50,000 bunch crossings can be saved and analyzed in detail. The job of the three–level trigger system is to select events of interest (such as those containing a Higgs boson), and to reject those containing more mundane interactions (such as low–energy inelastic $p\bar{p}$ scattering or di–jet production). Each event must pass each successive level of trigger, to be considered by the next level. Each later trigger level sees a smaller rate of events and has more time to process each event. Each level also has more information available than the previous one to use for making its decisions.
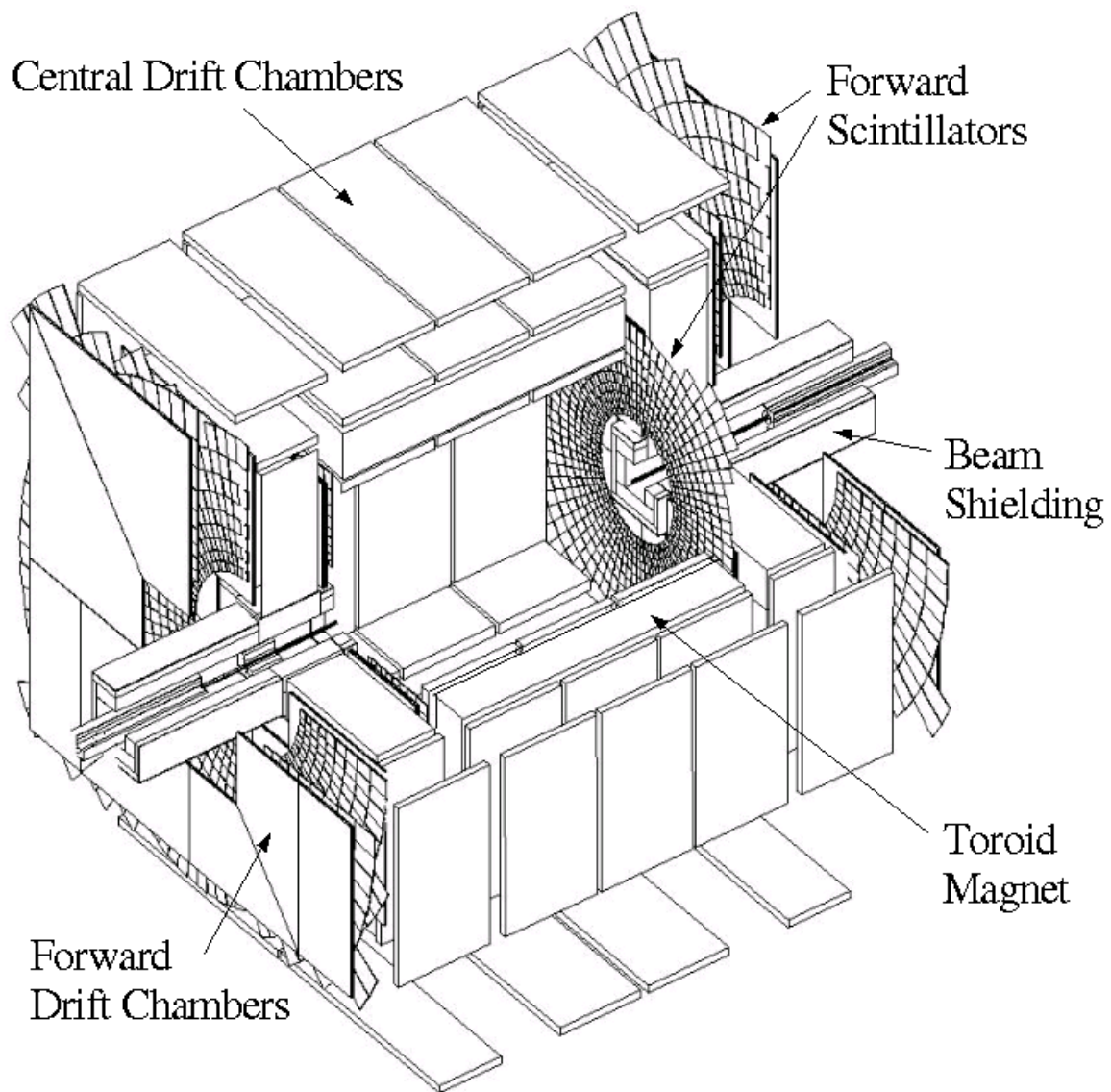
Figure 3.6: The muon system of the DØ detector, upgraded for Run II.

Level 1 and Level 2 Trigger Systems

The detector is able to trigger on calorimeter tower energies, high p_T tracks, missing E_T in the calorimeter, electrons, and muons at Level 1 (L1). This analysis only uses the calorimeter tower information in the first two levels of trigger decisions. Details of the trigger's calibration can be found in Appendix A. Efficiency determinations for simulated Higgs signals and backgrounds are discussed in Chapter 6.

A special, independent high–speed data path is used for triggering on energy in the calorimeter. A sum is made of the energy all the cells of in each η x ϕ = 0.2 x 0.2 calorimeter tower [10]. The tower energy sums are converted in analog to transverse energy, using specially designed resistor boards. These E_T sums are then 8–bit digitized and sent to electronics which compare the leading four E_T towers to programmed thresholds. If enough towers are above threshold, the L1 calorimeter trigger passes.

The rate of events passing this first calorimeter trigger level is further reduced by the Level 2 (L2) calorimeter trigger. The E_T measurements from all the calorimeter towers are used to perform a simplistic jet reconstruction. Starting with seed towers of E_T> 2.0 GeV, the E_T in 5x5 tower grids centered on the seed tower are calculated. Those which do not overlap and pass minimum E_T cuts (>5.0 GeV), are sent to the global Level 2 trigger processor. The global processor passes events based on whether the number of jets with E_T above given thresholds is sufficient. Events can also pass if the total E_T of jets is above a given threshold.

Level 3 Data Acquisition and Triggers

The rate of events passing all L2 trigger conditions, for all of the experiment's physics studies, is about 1 kHz. A factor of 20 rejection is needed at Level 3 (L3) to reduce the final event rate to 50 Hz. After a L2 trigger accept, the full data for the event

[10]The energy is also summed for the electromagnetic layers only, used for triggering on electrons.

is read out from each detector subsystem, and combined into a single computer for L3 trigger processing, a step called *data acquisition* (DAQ). If the event passes the L3 trigger, this data is recorded to tape. Details of the DAQ system can be found in Appendix B.

This analysis uses jets reconstructed at L3 to select events of interest. The jets are Run II legacy cone jets, very similar to those formed by the offline jet reconstruction algorithm (described in Chapter 4). Due to time constraints, which require each event to be processed in $\lesssim$100 ms, sacrifices are made in the complexity of the algorithms employed. The main deficiencies relative to the offline jet reconstruction algorithm are the lack of corrections for non–linear energy responses of calorimeter cell energies, the lack of suppression of noisier cells suffering from temporary read–out problems, and the lack of splitting or merging of jets from different seeds. Also, for the first version of the L3 trigger algorithm used, the jets' E_T was not corrected for the Z position of the primary interaction vertex. The loss of efficiency due to these effects is studied in Chapter 6.

Chapter 4

RECONSTRUCTION AND CALIBRATION

The detector data are reconstructed with sophisticated algorithms and carefully calibrated, to obtain accurate information about each event. This Chapter describes the steps which result in the high–level objects used to search for a neutral Higgs boson signal. The $p\bar{p}$ luminosity being delivered to the experiment is reconstructed from measurements made with the luminosity system. Measurements from the inner tracking detectors are clustered into *hits*, which indicate possible locations through which charged particles have passed. Pattern recognition is performed on the hits to reconstruct charged particle *tracks*. Tracks are used to find the $p\bar{p}$ interaction points (*primary vertices*), and to select the hard–scatter interaction point from any other inelastic $p\bar{p}$ collisions. Energy detected in the calorimeter is clustered into *jets*, which reflect the energy in quarks or gluons from the hard–scattering. The energy of each jet is calibrated, using the *jet energy scale*. Tracks associated with these jets are used by the *b–tagging* algorithm to search for evidence of bottom–quarks in the jets, as would often be present from the decay of neutral Higgs bosons.

Studies of the finding efficiency and quality of these objects are also presented. The energy resolution of the calibrated jets is crucial for identifying a Higgs boson resonance. Careful measurements are presented of the jet energy resolution in data and simulations. The performance of the bottom–quark jet (b–jet) identification (b–tagging) is also central to the neutral Higgs boson search. The b–tagging algorithm's characteristics are examined in detail.

4.1 Luminosity

During each beam crossing, the luminosity scintillators detect *coincident* particles originating from the interaction region, if an inelastic $p\bar{p}$ scattering event has occurred. The excellent timing resolution of the scintillators (0.2 ns) allows the collision point along the beam axis to be determined to about 6 cm. Excellent rejection from beam halo particles is achieved by requiring a reconstructed interaction point within the expected interaction region.

The instantaneous luminosity (for each combination of a proton and anti-proton bunch) is measured by counting the fraction of crossings with no $p\bar{p}$ coincidences detected. The probability of observing at least one inelastic collision, P(n>0), is given by:

$$P(n>0) = 1 - e^{-\mu} \tag{4.1}$$

μ is the average number of inelastic collisions per beam crossing:

$$\mu = \pounds \sigma_{eff}/r \tag{4.2}$$

$\pounds$ is the instantaneous luminosity, σ_{eff} is the effective inelastic $p\bar{p}$ cross–section (corrected for acceptance and efficiency), and r is the rate at which this bunch crossing occurs. Thus, the instantaneous luminosity is reconstructed to be:

$$\pounds = -(r/\sigma_{eff}) \ln(1 - P(n>0)) \tag{4.3}$$

The crossing rate, r, is very well measured, and equal to 7.58 MHz. The effective inelastic cross–section is the dominant source of uncertainty. It's value is approximately 43 mb, as measured at experiments at CERN and CDF during Run I. The value is also corrected for detector efficiency (about 91%) and acceptance (about 97%).

Each minute, approximately, a *luminosity block* is written to a database, recording the average luminosity for each bunch over the time period, as well as whether the data acquisition and detector hardware was performing properly.

4.2 Tracking

4.2.1 SMT Cluster and Hit Reconstruction

The signals from a group of adjacent silicon strips above threshold are combined into a *silicon cluster*. The center of the cluster is defined to be the charge–weighted average of the strip positions in the cluster, $\overline{n}$:

$$\overline{n} = \sum_i n_i \cdot c_i \tag{4.4}$$

where n_i is the position of the center of the ith strip in the cluster and c_i is its total collected charge.

The electrons and holes do not move directly towards the silicon strips, along the direction of the applied electric field, due to the presence of the 2T solenoidal magnetic field. Instead, they drift at an angle, the *Lorentz angle*, dependent on the strength of the magnetic field and the electron or hole *Hall-mobility*, the drift velocity in silicon. The Lorentz angle has been measured in test–beam data and in situ, by observing the likelihood of reconstructing single–strip clusters as a function of the track angle through the silicon wafer. The frequency of single–strip clusters is highest when the track angle is equal to the Lorentz angle, since all of the electrons and holes produced by the track drift along the same line. The Lorentz angle was measured to be about 4 degrees for the holes and 18 degrees for electrons, in good agreement with theoretical calculations and simulations [1]. The center of each silicon cluster is corrected for the average Lorentz angle.

---

[1]The holes are larger and have smaller Hall-mobility than the electrons.

The p–side and n–side of the silicon detectors have strips at relatives angles of 2 or 90 degrees, depending on their location in the SMT. Particles create clusters on both sides of the wafers, which can be combined, to determine the location along the strips where the particle has passed through. By using this "stereo" information, *silicon hits* are formed, representing a measurement of a point in space through which a particle is believed to have passed. The position of the hit can be reconstructed to as well as $\lesssim 10~\mu$m in the axial direction and $\lesssim 35~\mu$m in Z. The position in the third dimension is known to as well as 5 μm from alignment of the silicon wafers (described below).

4.2.2   CFT Hit Reconstruction

The adjacent CFT fibers above threshold in each doublet–layer are also grouped into *CFT clusters*. Most clusters are either singlet–clusters, containing just one fiber, or doublet–clusters, containing one fiber from each sub–layer of the doublet–layer. Clusters with larger numbers of fibers are also possible.

In each of the 8 CFT super–layers, consisting of two doublet–layers at a 3 degree relative angle, the overlapping CFT clusters are combined into *CFT hits*. The hits are measurements of points in space through which a particle may have passed. By using the geometry of the doublet–layers, hit resolutions are $\leq 100~\mu$m in the axial direction and about 2 cm in Z. The radial position of the hits are constrained by the positions of the axial fibers as measured through alignment (described below).

4.2.3   Track Pattern–Recognition

Pattern recognition is performed on the reconstructed hits, to determine a set of charged particle paths originating from near the interaction region [2]. The trajectories of charged particles in a perfectly solenoidal magnetic field are 3–dimensional helices.

---

[2]The track reconstruction software can be run in a special mode to find tracks from cosmic rays.

Small radial components of the magnetic field (from fringe effects near the ends of the solenoid) and scattering and energy loss through ionization of detector material slightly alter the charged particle paths, but in predictable ways [3]. An event with reconstructed tracks is shown in Figure 4.1. Since the view in the figure is parallel to the magnetic field lines, the paths of the tracks are seen as nearly circular arcs.

Two algorithms are used to perform pattern recognition: the Road Approach (GTR) and Histogramming Track Finding (HTF) [4]. Each algorithm creates a list of candidate tracks from the full set of reconstructed hits. If a hit is shared between a candidate track from each algorithm, it is assigned to the longer candidate track (the track with more hits assigned to it) and removed from the other candidate track. If the tracks have the same number of hits, the candidate track with a lower χ^2 fit of its path to its hits is chosen to keep the hit (and the hit is removed from the higher χ^2 candidate track). Once no remaining hits are shared, the candidate tracks with too few hits or fits too poor to their hits are removed. The final candidate tracks are then re-fit to their hits, using the Kalman smoothing technique, and the final track parameters and parameter errors are calculated.

GTR starts with track "stubs", curved paths through two hits which are consistent with coming from the interaction region. The two hits used to form each stub are usually required to be on the outer two layers of the CFT or the outer two layers of the SMT, but many special geometries are also allowed, to cover the overlapping regions between the CFT and the SMT disks, for instance. Each stub is propagated to additional tracking detector elements. Hits are added and the track is re-fit using a Kalman fitter if they are found within a search window. When more than one hit is found on a detector element within the search window of the track candidate, a new candidate is created for each possible new hit. If the χ^2/NDF of the track, a measure

---

[3]The magnetic field is surveyed using Hall probes. The location of the detector material and its density is very well mapped.

[4]A new algorithm, the Alternative Algorithm (AA), is used in later versions of the reconstruction software (versions p14 and above). It was not used to reconstruct the data for this analysis.

44
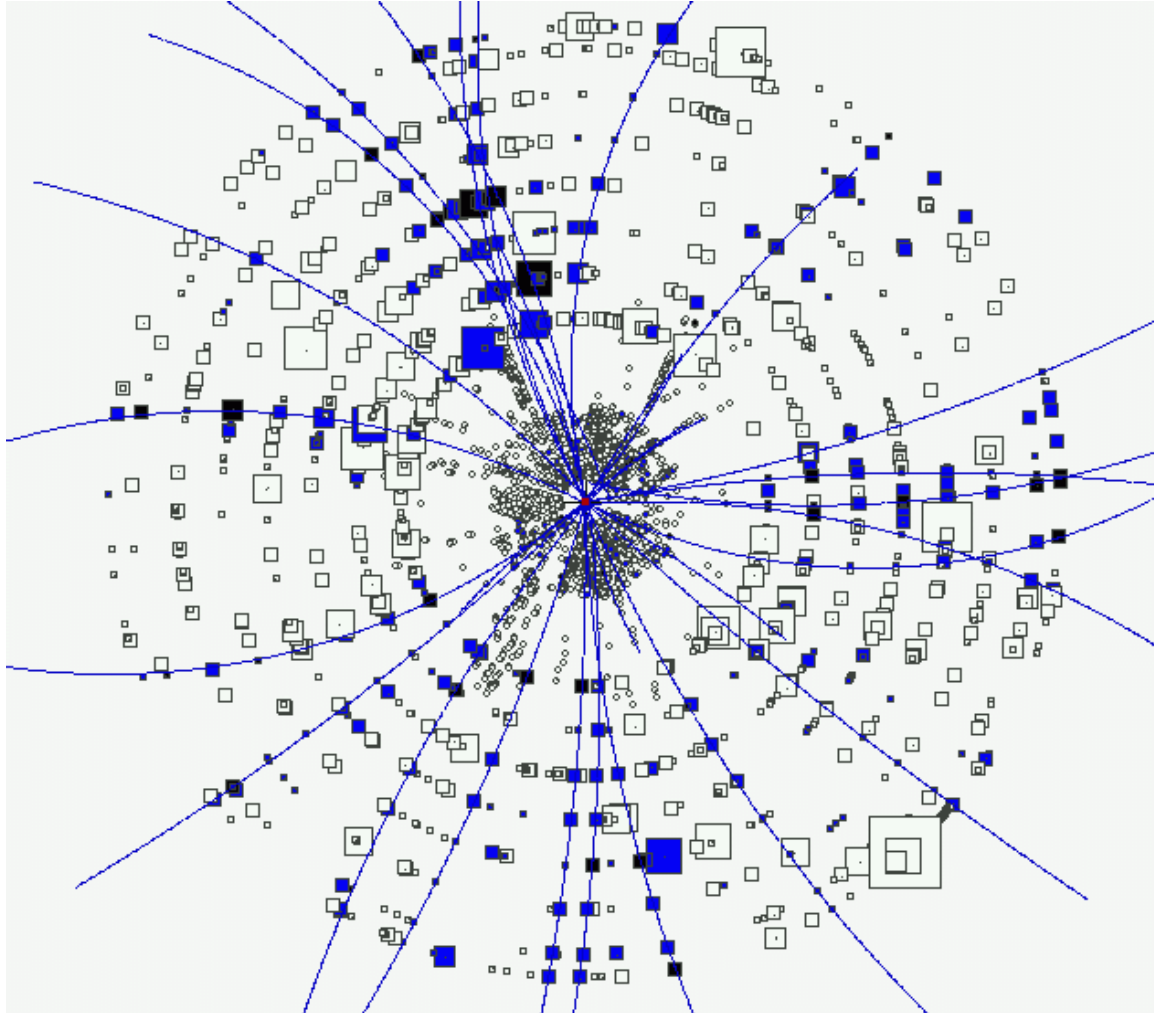
Figure 4.1: An axial view (looking down the beam-pipe) of an event recorded with the DØ inner trackers, showing the reconstructed hits and tracks. CFT hits are square and form the 8 layers on the outside. Silicon hits are drawn as small circles, and are innermost. Hits are colored solid if they are associated with a reconstructed track. The tracks are shown as solid lines, and are curved by the solenoidal magnetic field pointing out of the page.

of how consistent the hit positions are with the track parameters (taking all hit and track errors into account) grows above a threshold, the track candidate is discarded. Also, if more than one detector element is crossed without finding a reconstructed hit in the search window, the candidate is discarded.

HTF begins by forming track templates using the Hough transform technique. The Hough transform maps all hits on a circular arc passing through the interaction region to a single point, as shown in Figure 4.2. By looking for peaks in the histogram of Hough transformed hit data, the parameters for all circular arcs passing through the interaction region and four or more hits are found. Additional hits are added to these track templates using Kalman fitting techniques similar to those used in the GTR algorithm. Tracks with too many misses or too poor an overall fit are also discarded as new hits are searched for along the template tracks' trajectories.

4.2.4   Alignment

To match and combine the measurements of a single particle from the multitude of individual detector elements, the positions of the detectors must be known with the greatest possible accuracy. A high–precision measurement of where a particle has passed through a detector is useless unless the position of the detector itself is known to a high degree of accuracy. Care is taken during detector assembly and commissioning to place detector elements as near to their design positions as possible. They are then optically surveyed to determine their positions even more accurately. Each particle may traverse many detector elements, and follows a nearly smooth curve through space [5]. Thus, measurements from a single particle's curve can be used to constrain the positions of the detector elements. Using many particles, either from cosmic rays, or $p\overline{p}$ collisions, detector elements can be aligned, in situ. The process

---

[5]The particle's path is most greatly influenced by magnetic fields, which are well known and surveyed using Hall probes. The path is also deflected through scattering in material, whose density and position is also well measured through surveying.

46
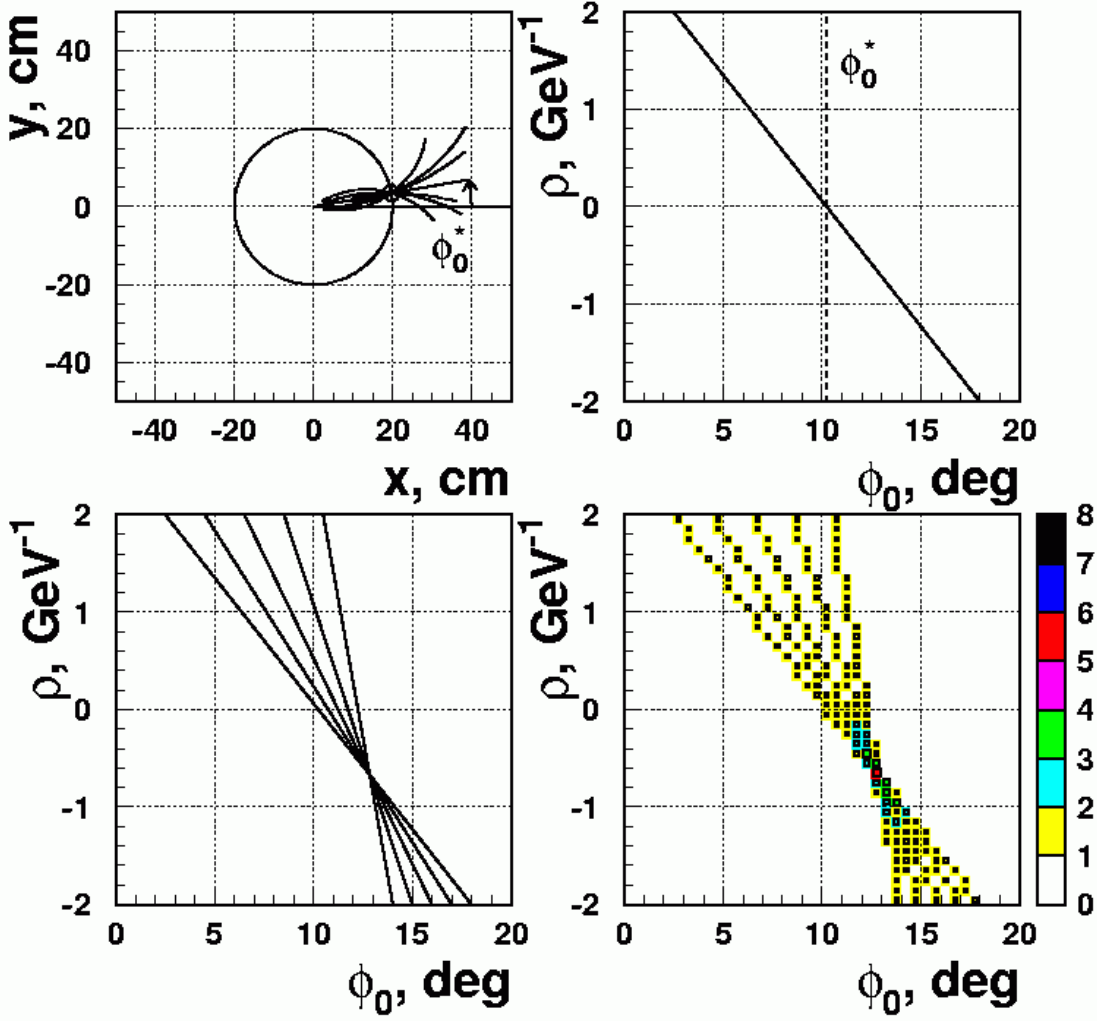
Figure 4.2: The Hough transform, used by the HTF track reconstruction algorithm, applied to a toy simulation of a single 1.5 GeV track coming from the center of the detector. The upper left plot shows the family of trajectories (circular arcs) through a given hit. The upper right plot shows the line in Hough transformed space corresponding to the locus of points which corresponds to each trajectory though the given hit. The lower left plot shows the family of lines corresponding to each of the five hits on the single 1.5 GeV simulated track. The lower right plot shows the histogram of the 5 Hough transformed hits. The peak in the histogram corresponds to the parameters (curvature and angle) of the original 1.5 GeV simulated track.

is iterative, first aligning larger detector pieces as rigid bodies, then aligning smaller detector elements of those detector pieces. For instance, first the SMT detector as a whole is aligned with the rest of the DØ detector, then the barrels are aligned relative to each other, then individual silicon wafers.

4.3   Primary Vertexing

Reconstructed tracks are used to find the locations of the primary vertices, points where $p\bar{p}$ inelastic collisions have occurred in the interaction region. First, all tracks in the event are fit to a common vertex (point in 3–dimensional space). If the χ^2/NDF of the vertex is >10, the track with highest contribution to the χ^2 of the vertex is removed from the vertex fit. This process is continued until either the vertex fit has a χ^2/NDF<10 or less than two tracks remain in the vertex fit. Vertex finding is repeated from the beginning, using all tracks not already associated with a primary vertex, until no more primary vertices are found.

For the range of instantaneous luminosities at which the data for this analysis was recorded ($1 - -4 \times 10^{30}$ cm^{-2}s^{-1}), about 0.5 inelastic $p\bar{p}$ collisions are expected in each event on average, in addition to the primary hard–scattering interaction. If more than one primary vertex is reconstructed in an event, the hard–scattering interaction is selected from amongst them. The tracks from an average inelastic $p\bar{p}$ collision, a *minimum bias* interaction, have smaller p_T on average than tracks from the hard–scattering interaction, as seen in Figure 4.3. Using minimum bias triggered data (only requiring a luminosity system scintillator coincidence), a probability is calculated for each track to be from a minimum bias interaction, based on its $log_{10}(p_T)$. The minimum bias probabilities for each track are combined into a probability for each of the primary vertices to be from a minimum bias interaction. The primary vertex with the lowest probability to be minimum bias is selected as the hard–scattering primary vertex. This vertex is used to determine each calorimeter tower's E_T from its energy,
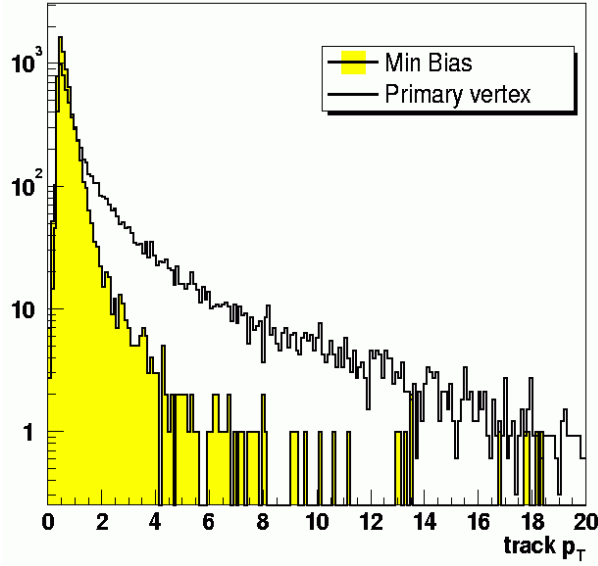
Figure 4.3: A comparison of the p_T of reconstructed tracks (in GeV/c) from simulated minimum bias interactions and hard–scattering events.

and is used for b–tagging (discussed below).

4.4 Jets

This search relies on the accurate reconstruction of jets to efficiently identify and study multi–jet final states. The methods used for reconstructing, identifying, and calibrating the jets are described in this Section. Jets are first reconstructed from the measurements made with the calorimeter. Then they are required to pass quality cuts, which reduces the fractions of fake jets (from calorimeter noise) and EM objects (photons and electrons). Jet energies are calibrated by correcting for detector and physical effects using the jet energy scale factors. The performance of the jet reconstruction, identification, and calibration is also studied.
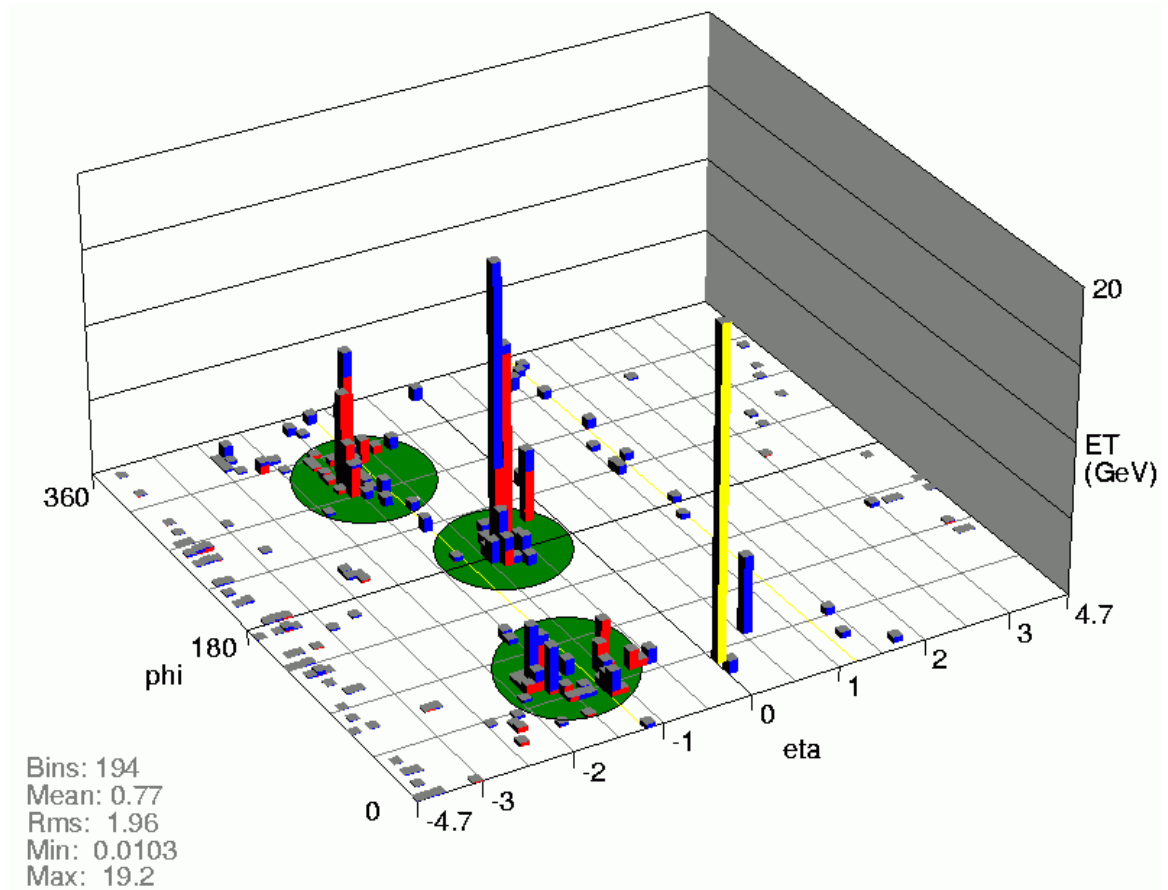
Figure 4.4: A recorded event from the DØ detector, showing the transverse energy measured in each calorimeter tower. Energy deposited in the EM layers of the calorimeter is shown in red. Energy in the hadronic layers is drawn on top of that in the EM and is shown in blue. Three jets have been reconstructed in the event and are circled in green. The missing E_T is shown in yellow, and is expected to be large since the energies of the jets have not yet been calibrated.

4.4.1   Jet Reconstruction

The Run II "improved legacy cone" algorithm [24] is used to reconstruct jets, both in data and in Monte Carlo. Figure 4.4 shows an event recorded from the DØ calorimeter with three reconstructed jets. Each tower's E_T is calculated from its total energy using θ', the polar angle between the beam axis and the tower center, as seen from the selected hard–scatter primary vertex [6]. "Seeds" are formed from all calorimeter towers with $E_T > 0.5$ GeV, and total $E_T > 1.0$ GeV within a cone of radius $\Delta R < 0.3$. The energy in a cone of radius of $\Delta R < 0.5$ around each seed is then calculated, and the E_T–weighted center of the cone is found [7]. This new center is then used as the cone axis, the energy in the cone of $\Delta R < 0.5$ is calculated again, and another new center is found. This procedure is iterated until a stable cone axis is found. Each stable cone found is defined as a *proto–jet*. In addition, the midpoints (in $\eta - -\phi$ space) between all pairs of proto–jets are used as seeds and iterated until a stable cone axis is found. Using the proto–jet midpoints as additional seeds makes the final reconstructed jets less sensitive to the initial seed positions [8]. Duplicate proto–jets (those with the same axes) are removed, as well as those with total $E_T < 8$ GeV. Then proto–jets are compared for overlapping regions. If the E_T contained in the overlapping region between two jets is greater than half of the E_T of either jet, the two jets are *merged*, and the jet energy and axis are recomputed. Otherwise, the jets are *split*, the towers are added to the nearest cone center only, and the energies and axes of each jet are recomputed.

---

[6] $E_T = E \cdot \sin \theta'$

[7] Other cone radii can be used, from $0.3 - 1.0$. This analysis uses a cone of radius 0.5.

[8] For instance, consider a parton shower that leaves most of its energy separated by a distance $0.5 < \Delta R < 1.0$. If no midpoint seed is used between the two high–E_T towers, the parton shower will most likely be reconstructed as two jets. It will be reconstructed as a single jet if enough soft-gluon-radiation (from NNLO) or calorimeter noise is present between the two high–E_T towers. Both of these effects are very hard to model accurately. However, using the midpoint as an additional seed always allows such a shower to be reconstructed as single jet. [25] [26]

4.4.2   Jet Selection Criteria

A set of quality cuts are applied to each reconstructed jet. These help to reduce fake jets from calorimeter noise [9]. EM objects such as photons and electrons are also removed. The following jet quality cuts are applied to each jet in the data and simulated samples:

- 0.05 <EM Fraction <0.95

  The EM fraction (EMF) is the fraction of transverse energy in the EM layers (1–7) of the calorimeter. Jets tend to deposit their energy uniformly in the EM and hadronic calorimeters, leading to an EMF of about 0.5 on average. Electrons and photons very rarely reach the hadronic layers of the calorimeter, and thus create showers with EMF >0.95.

- Coarse Hadronic Fraction (CHF) <0.4

  CHF is the fraction of transverse energy in the coarse hadronic layer of the calorimeter. Very few jets have a high CHF in the simulation. The coarse hadronic layer is known to be prone to noise.

- Hot Fraction <10

  The hot fraction is the ratio of transverse energy in the most energetic cell to that measured in the next most energetic cell.

- n90 >1

  n90 is the smallest number of cells required to contain 90% of the jet energy. If greater than 90% of the jet's energy is contained in a single tower, the jet is almost certainly reconstructed from a single hot calorimeter tower or cell.

---

[9]The taggability requirements, described below, further eliminate the fake jets.

- f90 $>0.8 - 0.5 \cdot$ CHF

  f90 is defined as n90/n_{item}, where n_{item} is the total number of calorimeter towers in the jet. This cut removes fake jets due to noise in the coarse hadronic layer of calorimeter, as discussed below in the Jet Selection Performance Section, and shown in Figure 4.9.

Jets are also required to have $|\eta|<2.5$ and $E_T>15$ GeV in order to be able to apply the jet energy scale (as described below). This is also the fiducial range for b–tagging.

4.4.3   Jet Reconstruction and Selection Performance

Jet finding and selection efficiencies are determined in Monte Carlo events and in data. A scale factor is derived between the data and MC efficiencies, as a function of jet E_T. The properties of jets are well modeled in the simulation, particularly above E_T of 15 GeV. The jet E_T spectrum for all reconstructed jets passing the jet quality cuts in data events, requiring only one jet with $E_T>95$ GeV, is compared to the spectrum in simulated events [10] in Figure 4.5. The jet E_T spectrum for jets failing the jet quality cuts in the same data and simulation samples is shown in Figure 4.6. The jet E_T spectra are very well modeled, for jets of $E_T>15$ GeV. At lower E_T (<15 GeV), there is significant loss of efficiency and many more jets failing the jet quality cuts in the data as compared to simulations. Overall efficiency for jet reconstruction and selection in the simulation rises from about 80% at a jet E_T of 15 GeV, the minimum considered in this analysis, to 95% at 25 GeV, and is nearly fully efficient by 40 GeV, as shown in Figure 4.7.

---

[10]The simulation uses the Pythia generator, with a p_T cut of 80 GeV, with events then run through the full detector simulation and reconstruction. A description of simulation methods is given in Chapter 5.
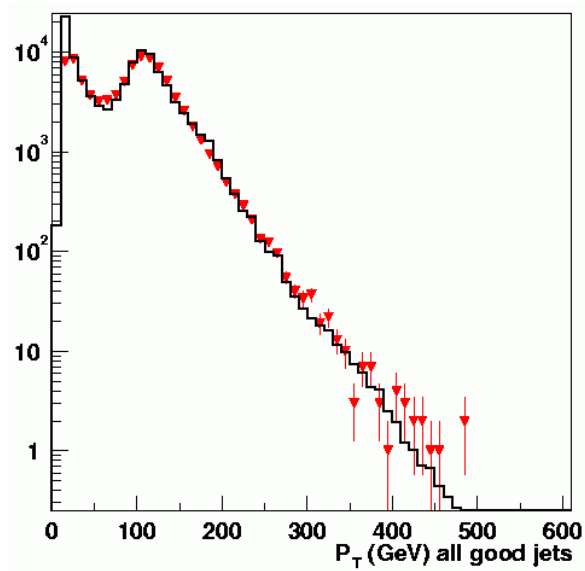
Figure 4.5: The E_T spectrum of all reconstructed jets passing the jet quality jets in data (points) triggered by requiring one jet with E_T>95 GeV, compared to simulation (solid histogram). (The simulation is Pythia events with a hard–scatter p_T cut of 80 GeV run through the full detector simulation and reconstruction).
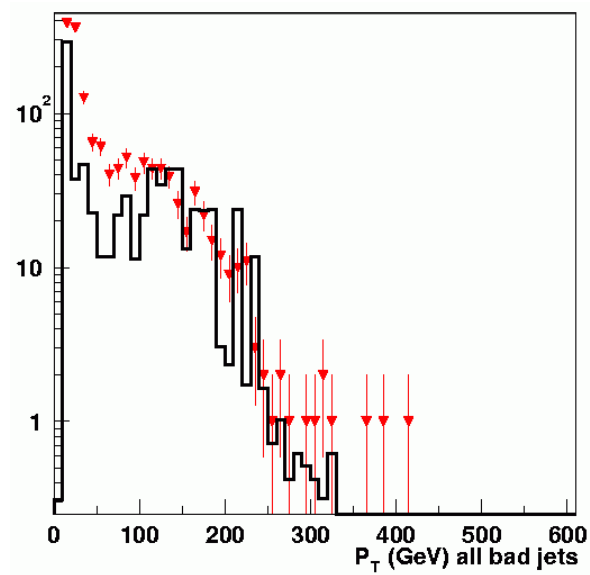
Figure 4.6: The E_T spectrum of all reconstructed jets failing the jet quality jets in data (points) triggered by requiring one jet with E_T>95 GeV, compared to simulation (solid histogram). (The simulation is Pythia events with a hard–scatter p_T cut of 80 GeV run through the full detector simulation and reconstruction).
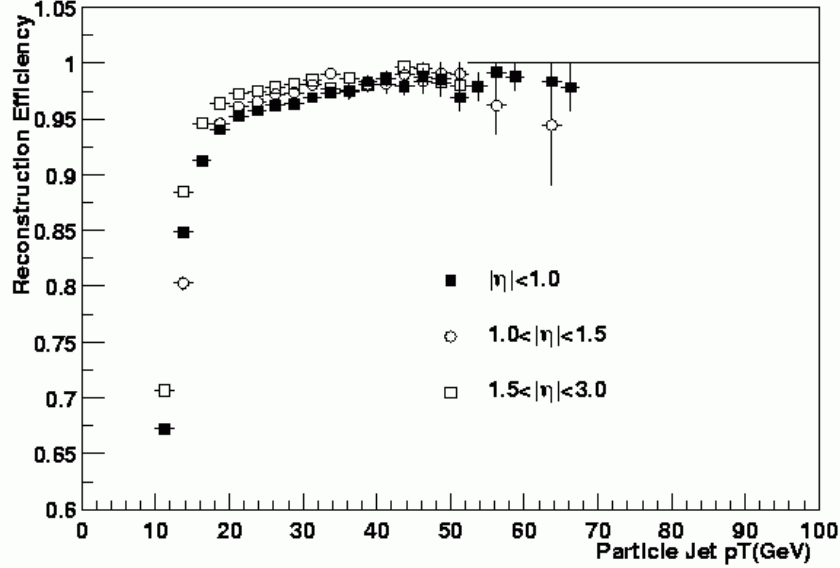
Figure 4.7: The efficiency for reconstructing and selecting jets in Monte Carlo events, as a function of jet E_T, for three difference $|\eta|$ ranges.

The jet reconstruction and selection efficiency data to Monte Carlo scale factor is shown in Figure 4.8. The identification of jets is slightly less efficient at lower E_T in data than in the simulation. The scale factor has been measured by using photon+jet events in data and comparing to Monte Carlo simulations of the same event type [41].

Fake jet rates can not be measured well in the simulation because calorimeter noise is not correctly modeled. However, fake jets tend to be dominated by one noisy tower, leading to a lower f90 value (fraction of towers containing 90% of the jet E_T), and higher CHF (fraction of jet E_T in the coarse hadronic layers), as seen in Figure 4.9. These fake jets are much more likely to fail the jet quality cuts, specifically the f90 requirement, f90 $>0.8 - 0.5 \cdot$ CHF (shown as a line in Figure 4.9). Jets in the simulation, where the calorimeter noise is lower, nearly all pass the f90 requirement. The fraction of fake jets remaining in the data after jet selection is determined by studying the f90 distribution, shown in Figure 4.10. The f90 distribution after all section cuts (except the f90 requirement) is fit to a sum of two Gaussian distributions,
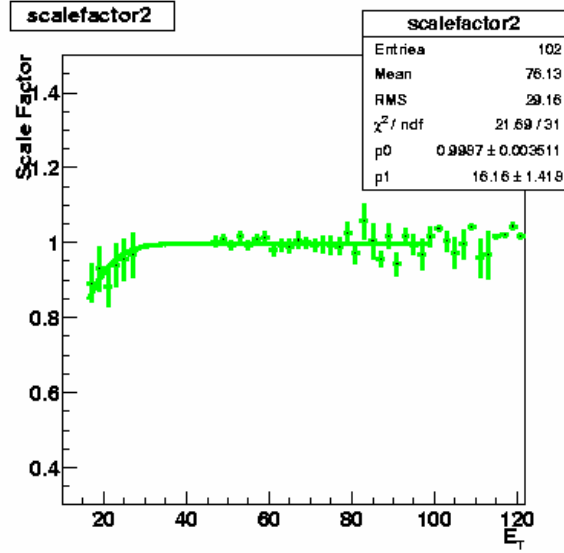
Figure 4.8: The data to Monte Carlo jet reconstruction and selection efficiency scale factor, as a function of jet E_T (calibrated using the jet energy scale).

representing good jets and fake jets. The fraction of fake jets is approximately 6%, as estimated from the number of jets in the fake Gaussian distribution extended into the region which would pass the f90 requirement. After requiring two tracks to be associated with the jet for taggability (as described below), even fewer fake jets remain ($<$1%), as derived in Chapter 5.

4.4.4   Jet Energy Scale

Jets passing the quality cuts are calibrated using the jet energy scale (JES). The calibration is a function of the total jet energy, the jet E_T, and the jet η as measured from the center of the detector (η_d). Jet energies are corrected according to:

$$E_{jet}^{corrected} = \frac{E_{jet}^{colorimeter} - E_{offset}}{R_{jet} \cdot R_{cone}} \qquad (4.5)$$
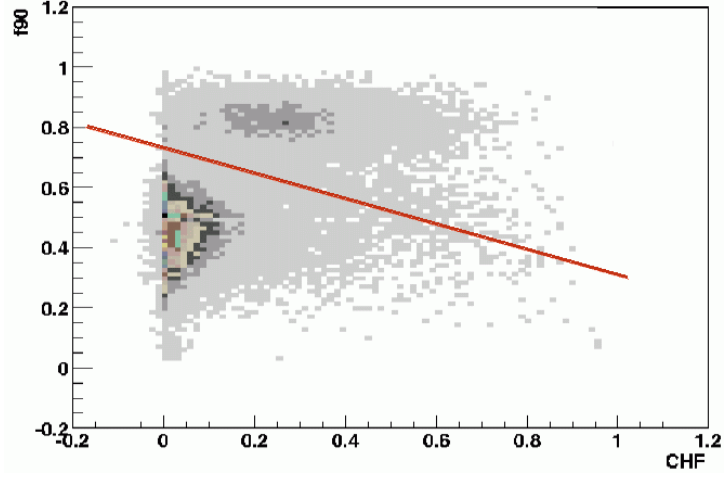
Figure 4.9: The f90 value of jets (the ratio of the number of towers required to contain 90% of the jet's energy to the total number of towers in the jet) compared to their coarse hadronic fractions (CHF), in a jet-triggered data sample. Fake jets tend to lie above the line representing the f90 jet quality cut (f90 $>0.8 - 0.5 \cdot$ CHF), and are thus removed by the jet selection process.
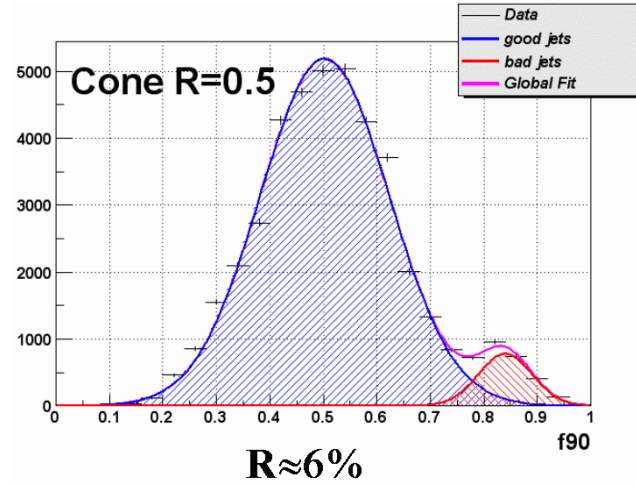
Figure 4.10: The f90 value of jets (the ratio of the number of towers required to contain 90% of the jet's energy to the total number of towers in the jet) in a jet-triggered data sample. The distribution of f90 is fit to a sum of two Gaussian distributions corresponding approximately to good and fake jets. The fraction of fake jets is approximately 6%, as estimated from the number of jets in the fake Gaussian distribution extended into the region which would pass the f90 requirement.

where $E_{jet}^{colorimeter}$ is the energy measured in the reconstructed jet cone, E_{offset} is the *offset energy* at the measured jet η_d, R_{jet} is the jet *response* correction factor, and R_{cone} is the fraction of the jet energy expected to be contained in the jet cone.

- Offset Energy

  Inelastic $p\bar{p}$ collisions in the bunch crossing other than the hard–scatter collision (minimum bias interactions) and calorimeter noise add energy to calorimeter towers, even in the absence of jets. This energy, a function of η_d, must be subtracted from each reconstructed jet's transverse energy. The offset energy is shown in Figure 4.11, as measured in data triggered requiring only an inelastic $p\bar{p}$ interaction (minimum bias data). The curve corresponding to the medium luminosity is used as the central value in the offset energy correction. The difference in the offset energy between the other two instantaneous luminosities is used to estimate the systematic error involved. An identical plot is made for simulated minimum bias events and used to correct for offset energy for jets in Monte Carlo events.

- Jet Response

  The response of the calorimeter is the ratio of the measured energy to the true energy deposited in the calorimeter. The response varies as a function of jet energy, due to non–linearities, and as a function of the η_d where the jet is measured in the calorimeter, due to varying geometry. Figure 4.12 shows the jet response in data as a function of the jet energy. Corrections are obtained by requiring E_T balance in photon+jet events in data, and also separately in Monte Carlo. These events have the advantage that the photon E_T is very well measured (compared to jet E_T). Also, the absolute scale of the photon energies are known, by calibrating the EM energy scale using $Z{\rightarrow}e^+e^-$ events. The electron and positron energies in the $Z{\rightarrow}e^+e^-$ events are calibrated with the
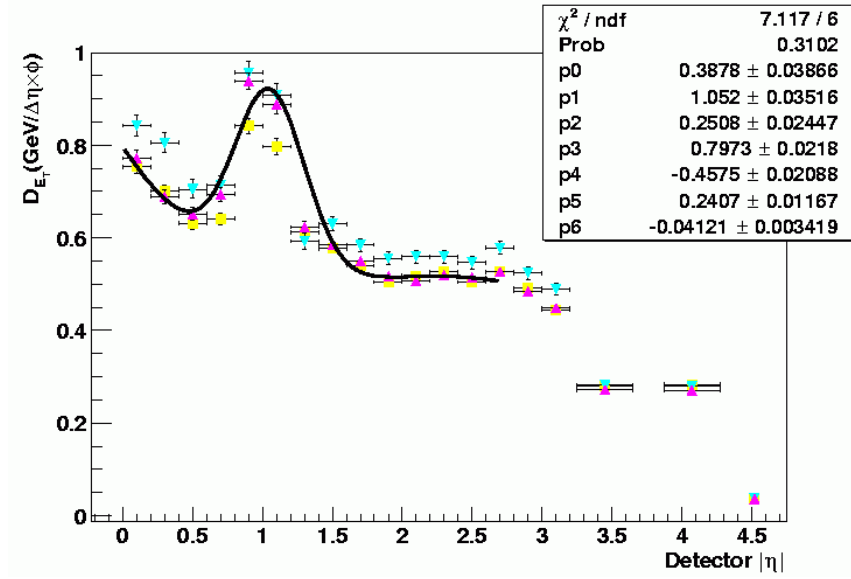
Figure 4.11: The offset energy in the calorimeter, as a function of η_d as measured in minimum bias triggered data for three instantaneous luminosities, approximately 1, 2, and 3×10^{30} cm^{-2}s^{-1}. The curve corresponding to the medium luminosity is used as the central value in the offset energy correction. The difference in the offset energy between the other two instantaneous luminosities is used to estimate the systematic error involved.
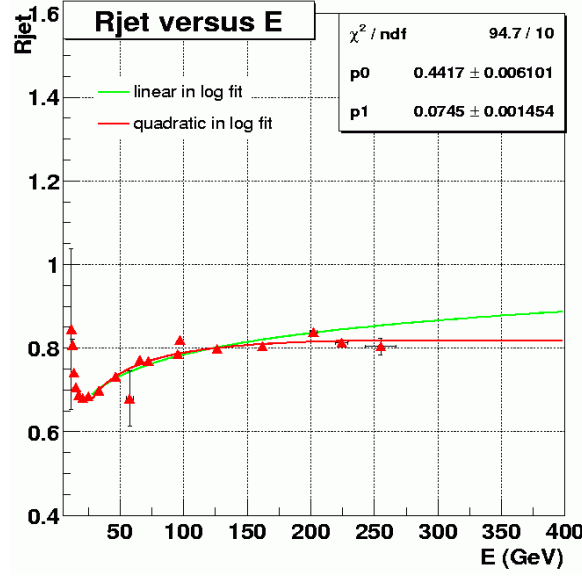
Figure 4.12: The energy response of the calorimeter in data as a function of jet energy.

tracks' momenta as reconstructed from the inner tracking detectors.

- Out of Cone Showering

  Some of a jet's energy is typically radiated outside of the cone area and is not included in the measured jet energy. The energy in cones of radius from 0.1 to 1.5 in η x ϕ space around the reconstructed jet axis is measured. This distribution is called the shower *profile*, and is shown for central jets ($\eta<0.7$) in photon+jet data in Figure 4.13. In the central region, 92% of the jet energy is included in a 0.5 radius cone on average in the data.

- Muon Energy

  Jet energies are corrected for energy lost through muons when they are identified within the jet cone radius. Muons leave only about 3 GeV of their total E_T in the calorimeter on average. The energy expected from the associated neutrino(s) from the muon's production is also added back to the jet's energy. b–jets con-
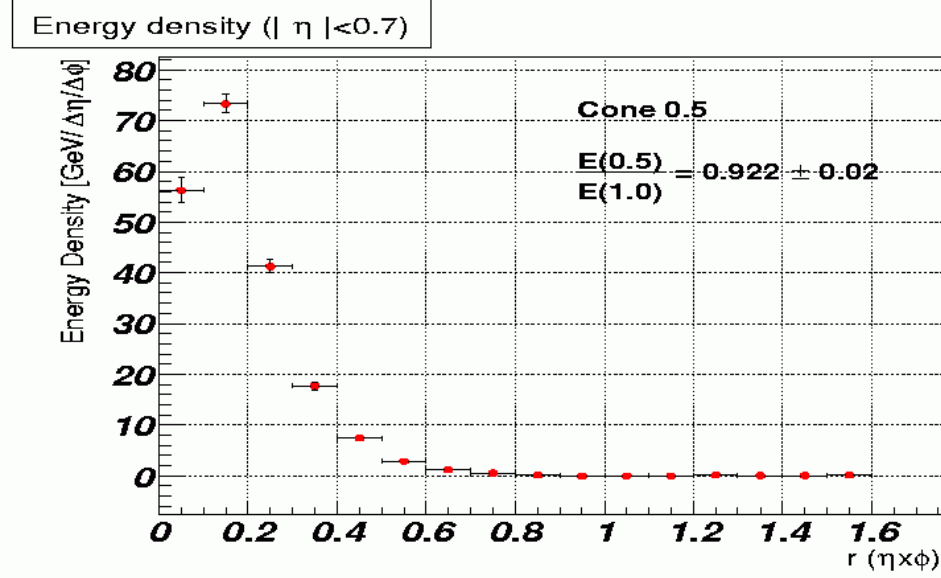
Figure 4.13: The jet showering profile for central jets in photon+jet data. On average, 92% of the jet energy is included in a cone of radius $\Delta R < 0.5$.

taining muons are studied in simulated events, where the true (particle–level) jet energies are known. The fraction of the measured muon energy which should be added back to the jet energy on average is calculated, as shown in Figure 4.14.

Plots of the total jet energy scale corrections and their uncertainties as a function of E_T and η are shown in Figures 4.15 and 4.16. Typical relative jet energy scale correction values are +30% for data and +20% for MC jets. The total uncertainty in the jet energy scale corrections is about 5% (mainly systematic). The major sources of uncertainty arise from effects at different instantaneous luminosities, variations from tighter primary vertex requirements, and changes in energy balance from different $\Delta\phi$ cuts between the photon and the jet in each event.
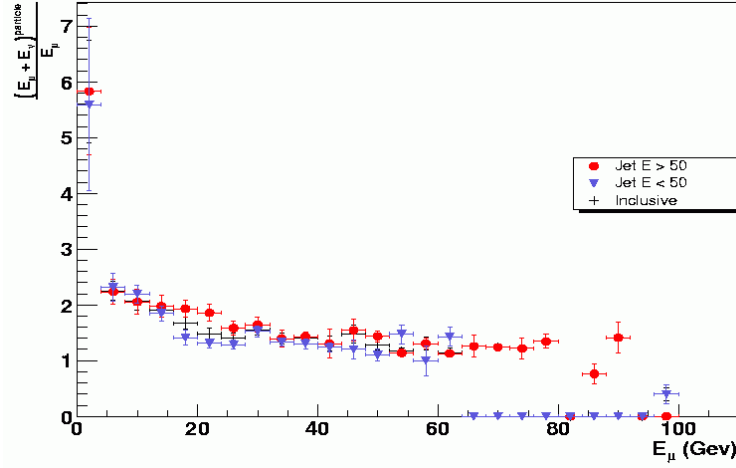
Figure 4.14: The jet correction factor for measured muon energy associated with a jet. The fraction of the muon energy to add back to jet is derived from simulated b–jets containing muons. The correction factor is shown for low (<50 GeV) and high (>50 GeV) energy jets and the inclusive jet sample.

4.4.5 Di–jet Mass Resolution

The invariant mass of the leading two jets, which have been corrected back to the particle level by the JES corrections, is given by:

$$M_{j_1 j_2} = 2 \cdot E_{T1} E_{T2} \cdot (\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2)) \tag{4.6}$$

The invariant mass distribution is peaked near the mass of the Higgs boson (m_h) for the Higgs boson signal. An optimized Gaussian fit to these mass peaks (dependent on m_h) is used to isolate the Higgs signal region, for normalizing the backgrounds. Thus the di–jet mass resolution, or equivalently the jet energy resolution, must be properly understood in the data. In particular, the differences in jet energy resolution between the data and Monte Carlo are important, since the signal shapes are derived from Monte Carlo and then used for the search in the data.

Two independent data samples are used to measure the jet energy resolution. The
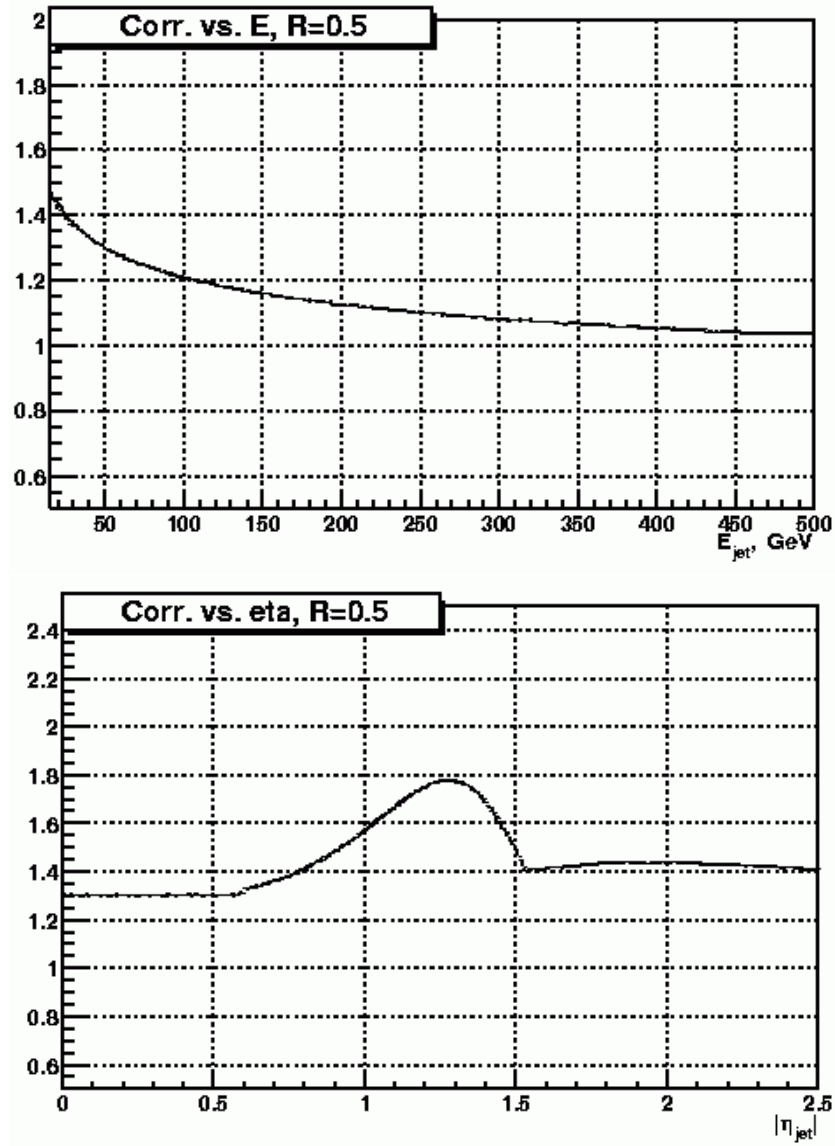
Figure 4.15: The jet energy scale correction factor measured for jets in data as a function of E_T (top) and η (bottom).
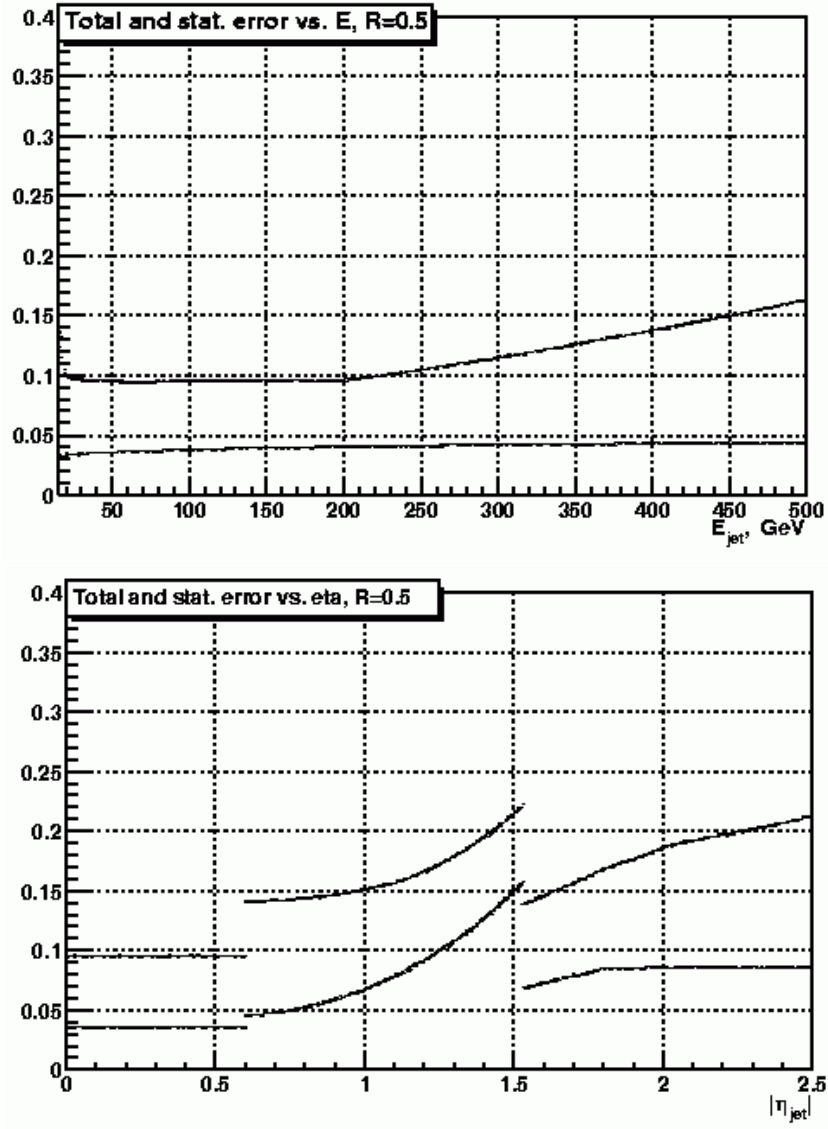
Figure 4.16: The jet energy scale correction uncertainties (statistical and total) measured for jets in data as a function of E_T (top) and η (bottom).
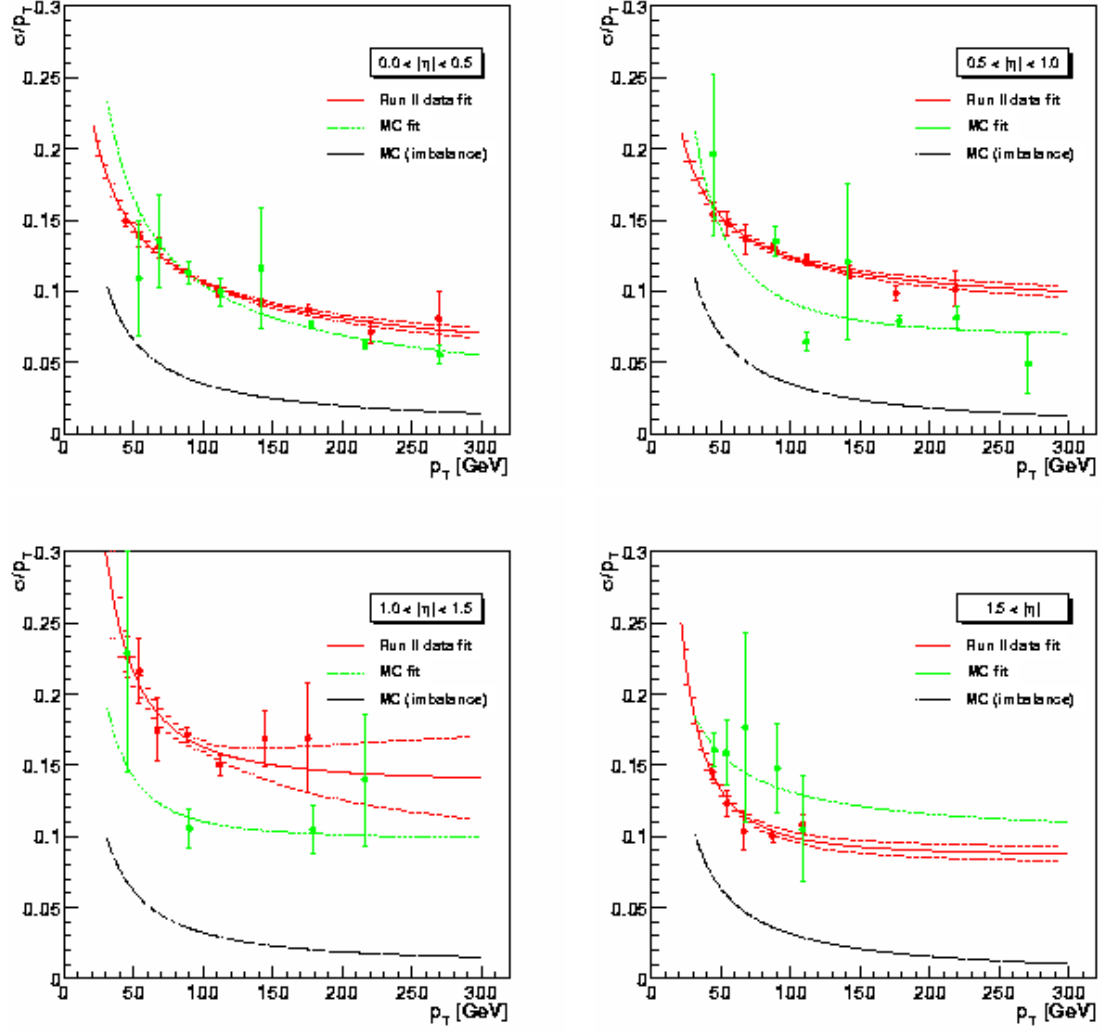
Figure 4.17: Final jet E_T resolutions in different η ranges for data (red) and Monte Carlo (green). The black curve indicate the Monte Carlo particle–level E_T imbalance, which is the best resolution theoretically achievable.

first uses di–jet events and studies the E_T balance between the two leading jets. The data sample used is a combination of single–jet triggered events with E_T thresholds of 25, 45, 65, and 95 GeV corresponding to about 10 pb^{-1}. The Monte Carlo samples are di–jet events, with E_T thresholds of 20–160 GeV, generated with Pythia v6.202. The jets in each event are required to be back–to–back, by imposing the requirement $\Delta\varphi > 175$ degrees. This cut helps to reduce the contamination from events where one of the jets has radiated a hard gluon.

The jet energy resolution is roughly $\sqrt{2}$ times the E_T imbalance in each event, as derived from Equation 4.6. Two corrections are made to derive the jet energy resolution more accurately from the energy imbalance in di–jet events. A cut is placed on the E_T of the third jet in the events. Events with higher E_T third jets will give larger E_T imbalance, which is calculated only using the two leading jets in the event. By varying the cut on the third jet E_T and observing the effect on the E_T imbalance, a linear relation between the two can be obtained. Extrapolating to a third jet E_T cut of 0 GeV, the best estimate of the true jet energy resolution, devoid of soft–radiation effects, is achieved.

The other applied correction accounts for the fact that the E_T, calculated from the two leading jets in di–jet events, is not truly balanced in di–jet events. Some energy is radiated outside the jet cones and is not perfectly corrected for by the JES corrections. The magnitude of this effect, as a function of jet E_T, is calculated in Monte Carlo. This unavoidable jet E_T imbalance is then subtracted in quadrature from the imbalance measured in data events.

The comparison between Monte Carlo jet energy resolutions and those in di–jet data events, as a function of corrected jet E_T and in different jet η bins, is shown in Figure 4.17. In general, the jet energy resolution is well modeled by the Monte Carlo. For jets of energy typical to those expected from the Higgs, with E_T of about 60 GeV, the resolution is about 12%, except in the Inter-Cryostat Region (ICR) of the calorimeter, where it rises to 18%.
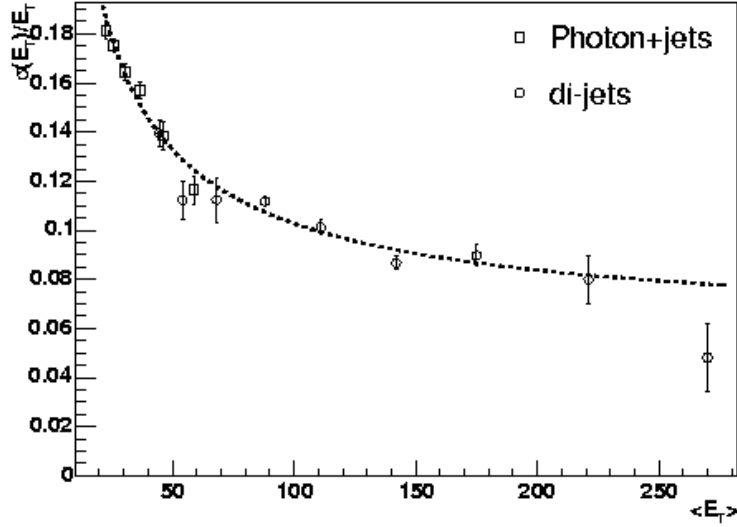
Figure 4.18: The jet energy resolution extended to low E_T using photon+jet data.

The other data sample type used to study jet energy resolution is photon+jet events. Again, the imbalance in E_T is used to derive the jet energy resolution. The photon energy resolution is much better than the jet energy resolution. The jet energy resolution is again corrected for the soft–radiation effects and MC–level particle imbalances. The advantage of this second sample is that it extends to much lower jet E_T values, since the EM trigger is efficient at much lower E_T than the jet triggers. The jet energy resolutions, after both corrections, from both the di–jet and photon+jet samples, are shown in Figure 4.18. The jet resolutions calculated from both samples agree at medium E_T, adding confidence to the extrapolation to lower E_T with the photon+jet data.

4.5   b–Tagging

The pure and efficient identification of b–jets in multi–jet final states is critical to this analysis. Events containing multiple b–jets must be selected from the large number
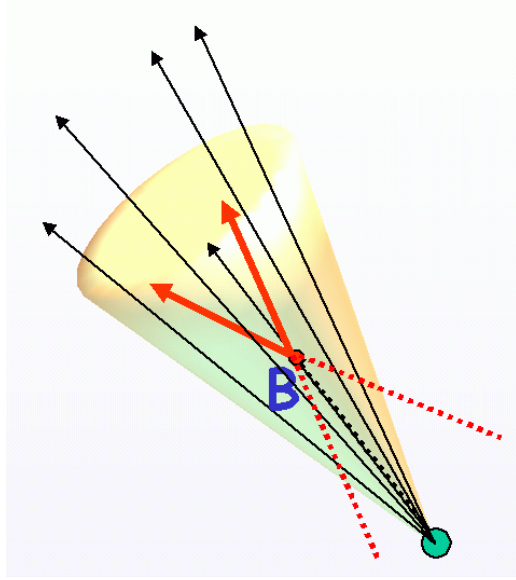
Figure 4.19: A drawing which shows tracks associated with a jet, and a secondary vertex reconstructed at the decay point of a B–hadron from the large impact parameter tracks from the decay.

of (mostly light–quark and gluon jet) multi–jet events. Accurately understanding the kinematical biases introduced by the b–tagging is also necessary. Differences between the b–tagging in multi–jet states as opposed to simpler event topologies are understood. The differences in b–tagging characteristics between data and Monte Carlo are also studied.

The b–tagging algorithm used relies on the fact that b–hadrons decay through the Weak force after traveling an average of about 3mm (at a p_T of 50 GeV/c) into an average of about 5 charged particles, along with other neutral particles (see Figure 4.19). These charged particles tend to have high p_T(>0.5 GeV/c) as well as a large impact parameter (IP), the closest distance that the particle would come to the primary interaction point along its trajectory. The algorithm attempts to identify the location of the *secondary vertex*, where the b–hadron (or sometimes a daughter c–hadron) decayed by finding a common point where the large impact parameter tracks
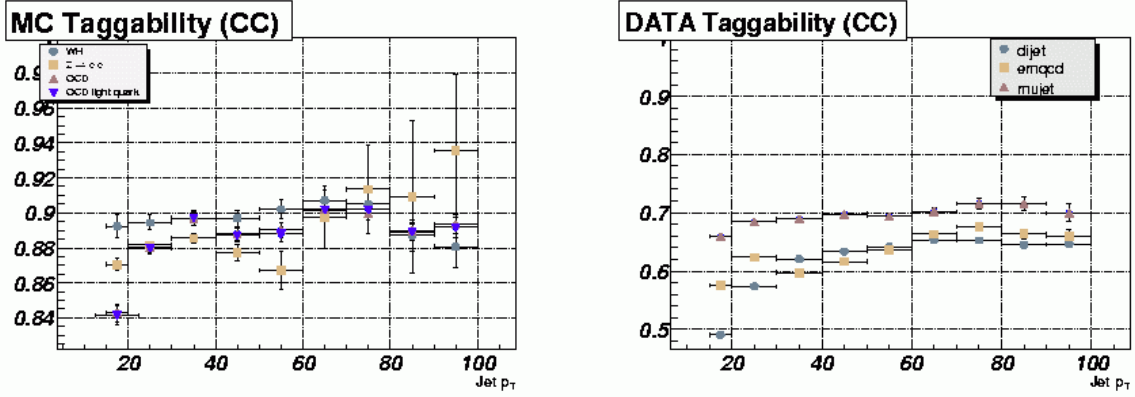
Figure 4.20: The taggability of jets in data (right) and Monte Carlo (left) events in various samples, in the central region ($|\eta|<1.2$) as a function of jet E_T.

intersect.

4.5.1 Taggability

No algorithm can identify a b–jet if it does not meet certain minimal requirements. A jet meeting these requirements is called *taggable*. Since this analysis attempts to tag each jet in the multi–jet final state, each jet is first required to be taggable. Requiring each jet to be taggable also significantly reduces the fake jet rate, as mentioned above, and as will be derived in Chapter 5. Each jet is required to have at least two associated tracks with $p_T>0.5$ GeV/c, within $\Delta R<0.5$ of the jet axis, and ≥ 3 SMT hits and ≥ 7 (0) CFT hits in the central (forward) region to be taggable.

Also, each event is required to contain a primary vertex reconstructed with ≥ 4 tracks attached to it, in order for any of the jets in that event to be taggable. Less than 2% of events passing initial data selection cuts (described in Chapter 5) failed to have a primary vertex with ≥ 4 tracks.

Approximately 75% of jets are taggable in data and 85% in Monte Carlo. This difference is taken into account in detail by the b–tagging data to MC scale factor (described below). The taggability in data and MC in the central region ($|\eta|<1.2$)

are shown in Figure 4.20 as functions of jet E_T for various samples. The taggability is very similar for all of the samples, and differences are taken as contributions to the b–tagging systematic error.

4.5.2 Secondary Vertex Tagger (SVT) Algorithm

The algorithm used to identify jets containing bottom–quarks is the Secondary Vertex Tagger (SVT) [30] [31]. SVT identifies b–jets by first reconstructing all secondary vertices in each event. The secondary vertices are then subjected to selection criteria, which removes many secondary vertices not from the decay of a heavy (c or b) quark. Jets are b–tagged if they can be matched to a passing secondary vertex. The Extra Loose vertex definition is used, which has the highest b–tagging efficiency, but also a larger fake rate (efficiency to tag a light–quark or gluon jet). Other definitions have more stringent requirements on the tracks used to reconstruct the secondary vertices.

The reconstruction of secondary vertices begins by finding *track–based jets*. The tracks are clustered using a 3–dimensional cone[11] algorithm of radius 0.5. Tracks are each required to have $\geq$ 2 SMT hits, $p_T >$ 0.5, IP$<$ 0.15 cm, and Z–IP$<$ 0.4 cm [12]. Tracks within the track–based jets are then selected which have large transverse (2–dimensional IP in the plane transverse to the beam-line) IP significance (≥ 3.0)[13]. All two–track "vertex seeds" (vertices formed from two tracks with a χ^2 probability ≤ 100) are found from the selected tracks in each track-jet. Additional tracks in the track–based jets pointing to the seeds are then attached to each seed if the resulting

---

[11] The fact that the cone algorithm is 3–dimensional means that tracks with widely different Z positions along the beam–axis ($\gtrsim$1 cm) are not clustered with each other. This helps to decrease the confusion of the algorithm by additional minimum bias interactions.

[12] The Z–IP is the impact parameter projected onto the Z axis: the closest the track comes in the Z direction to the hard–scatter primary vertex.

[13] The "significance" of a quantity is its measurement divided by its mean deviation (its "error"). So the IP significance is the IP divided by the track's IP error, calculated by projecting the track's error matrix along the direction parallel to the vector pointing from the primary vertex to the point on the track of closest approach to the primary vertex.

χ^2 contribution to the vertex fit is ≤ 15. A track is allowed to be attached to more than one vertex seed, so the presence of fake vertices does not impact vertex finding efficiency. The resulting reconstructed secondary vertices must then pass the following selection criteria:

- The "2–dimensional decay length", $|\overrightarrow{L}_{xy}|$, of the secondary vertex must be ≤ 2.6 cm. $\overrightarrow{L}_{xy}$ is the vector from the hard–scatter primary vertex to the secondary vertex, projected transverse to the beam-line.

- The "collinearity" of the secondary vertex must be ≥ 0.9. Collinearity is defined as the inner–product of $\overrightarrow{L}_{xy}$ with $\overrightarrow{p}_{xy}$, the momentum of the secondary vertex, defined as the sum of the transverse momentum of all its attached tracks. The collinearity requirement selects secondary vertices consistent with traveling from the hard–scatter primary vertex, since their momentum is pointing in their direction of travel.

- The secondary vertex must have ≥ 2 attached tracks.

- The decay length significance, $|\overrightarrow{L}_{xy}|/\sigma(|\overrightarrow{L}_{xy}|)$, of the secondary vertex must be ≥ 5, where $\sigma(|\overrightarrow{L}_{xy}|)$ is the error on the decay length of the secondary vertex, as determined from the vertex fit based on the reconstructed tracks' estimated errors.

A procedure called V_0 *removal* is performed to reduce the number of jets that are b–tagged which contain long–lived particles other than b–hadrons [33]. Two–track secondary vertices are found which are consistent with being a K_s^0, Λ, or a photon conversion into an e^+e^- pair. The procedure is based on the angle between the two tracks and their invariant mass, which can be compared to the known masses of common light–hadrons (or 0 mass and angle for photon conversions). These secondary vertices are removed from the candidates used to b–tag jets.

The remaining secondary vertices are matched to calorimeter jets if the ΔR between the vertex and the jet is less than 0.5. A calorimeter jet is tagged as b–jet if it has at least one matched secondary vertex.

4.5.3 Performance in Data

Measuring b–tagging performance in data is very difficult due to the impossibility of obtaining a pure b–jet or light–jet sample. It is also difficult to obtain a large sample of events with an enhanced fraction of b–jets. A large sample of b–jets is needed to measure the b–tagging efficiency as a function of various kinematic variables. Studies of b–tagging efficiency use primarily the μ–in–jet sample, which requires a muon with $p_T > 6$ GeV/c matched to a reconstructed jet within $\Delta R < 0.5$. The μ–in–jet sample has an enhanced fraction of b–jets, since b–quarks decay about 20% of the time (including both direct and cascade decays) to a high–p_T muon. Fake rate studies use jet–triggered and EM–triggered data, with at least two reconstructed jets.

Four methods are used to study the b–tagging efficiency in data. They agree well, within statistical errors, as shown in Table 4.1. The variation in the b-tagging efficiency between the methods is used to derive the systematic uncertainty involved. Relative uncertainties of 5% in the central region and 10% in the forward region are assigned.

The first method uses a fit to the p_T^{rel} distribution of the muons in jets to templates, derived in MC, before and after b–tagging. The p_T^{rel} is the p_T of the muon relative to the combined jet plus muon momentum axis. Since light–quark jets have muons whose p_T^{rel} values are smaller than for b–jets (due to the higher mass of the b–quark), a fit indicates approximately what fraction of those jets contain b–hadrons. Separate templates are derived for muons with $p_T > 10$ GeV/c, and for those with $p_T < 10$ GeV/c. Figure 4.21 shows a p_T^{rel} fit before and after requiring a b–tag to be present in each event. The total number of events remaining after b–tagging and the amount by which the fraction of b–quarks increases after requiring a b–tagged jet are used to
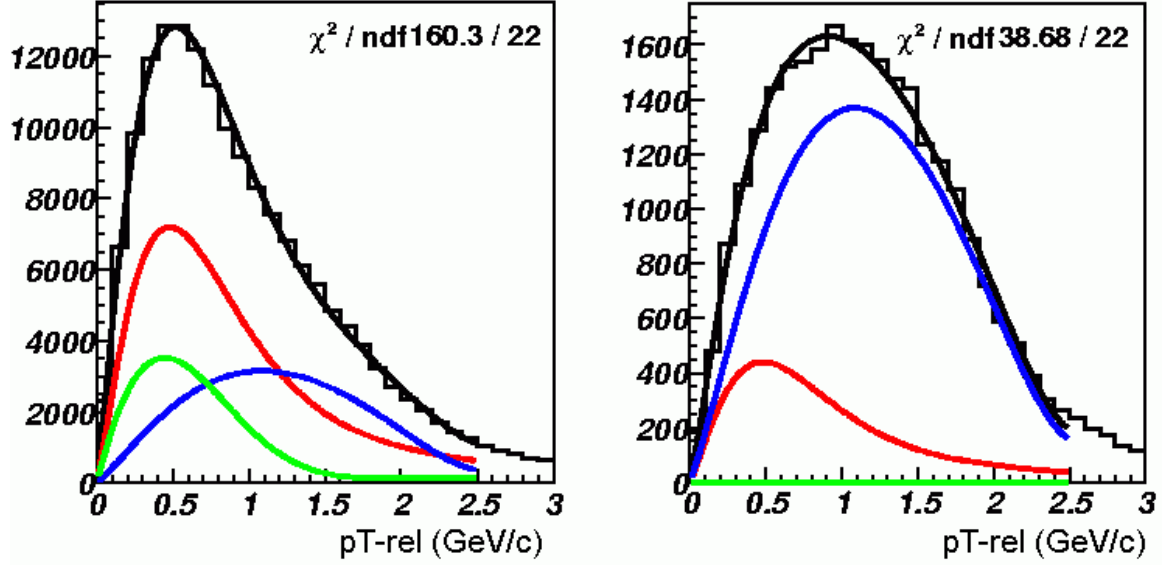
Figure 4.21: A fit of the p_T^{rel} distribution of muons in jets in the μ–in–jet sample to a sum of the light–quark and gluon (green), c–quark (red), and b–jet (blue) muon p_T^{rel} templates before (left) and after (right) requiring a b–tagged jet in each event. The χ^2/NDF values indicate the quality of the fit of the data to the templates. The fit shown is for muons with $p_T>$10 GeV/c only.

Table 4.1: SVT b–jet and light–jet b–tagging efficiencies for taggable jets measured in a variety of ways in data events, and their averages, in the central and forward regions.

Method	Efficiency (%) ($|\eta|<1.2$)	Efficiency (%) ($|\eta|>1.5$)
Single–μ p_T^{rel} fit	55.2±5.8	46.1±11.6
Away–jet b–tagged	54.6±2.0	35.9±8.4
SystemD	51.5±2.6	33.4±8.3
Di–μ ($p_T>$1GeV/c)	65.5±18.1	–
Average b–tag eff.	53.6±1.5(stat)±2.7(sys)	37.0±5.2(stat)±5.6(sys)
Jet data (negative)	2.0±0.04	1.67±0.06
EM data (negative)	1.68±0.04	1.60±0.05
Jet data (positive)	2.05±0.08	1.60±0.10
Average light–tag eff.	1.86±0.03(stat)±0.19(sys)	1.64±0.06(stat)±0.25(sys)

measure the b–tagging efficiency.

The away–jet method [34], studies the ratios of single–tagged to double–tagged events, again using the p_T^{rel} fits to determine the fraction of b–jets in the single b–tagged data sub–sample.

The third method, SystemD, does not depend on the p_T^{rel} fits. Instead, two data samples with different b–jet fractions (the μ–in–jet sample and an opposite tag sub–sample) and two taggers (a lifetime tagger and the requirement of a muon with $p_T^{rel}>0.7$ GeV/c) are used [35]. Solving a system of 8 (sometimes non-linear) equations derived from the number of jets b–tagged in each sub–sample gives results for both light–jet and b–jet tagging efficiency. The cut on the p_T^{rel} of the muon can be varied to evaluate the systematic error of the measurement.

The last method uses di–muon–in–jet events, where each muon has a $p_T^{rel}>1$ GeV/c. MC shows that such a sample has nearly 99% b–jet purity. However, the sample has very low statistics, due to the rarity of the process. Counting the number of events which are either single or double b–tagged in the sample leads directly to a measurement of the b–tagging efficiency.

Two methods are used to measure the b–tagging fake rates (efficiency for b–tagging a jet which contains no c– or b–hadron). The first uses *negative tags*, jets with vertices reconstructed behind the primary vertex (with respect to the jet axis) but which otherwise would have been matched to the jet [14]. Monte Carlo is used to estimate the correction factor (from remaining V^0's, etc.) for the number of positive tags that would be produced given the resolution obtained from the measured negative tag rate. Negative tag rates are measured in the jet–triggered data as well as the EM–triggered data, and the results are compared (see Table 4.1).

The other method uses the positive tag rate in jet–triggered events, but corrects for the c– and b–jet fractions that contribute to the positive tag rate using Monte

---

[14]For a negative tag, the ΔR between the secondary vertex and the negative jet axis is < 0.5.
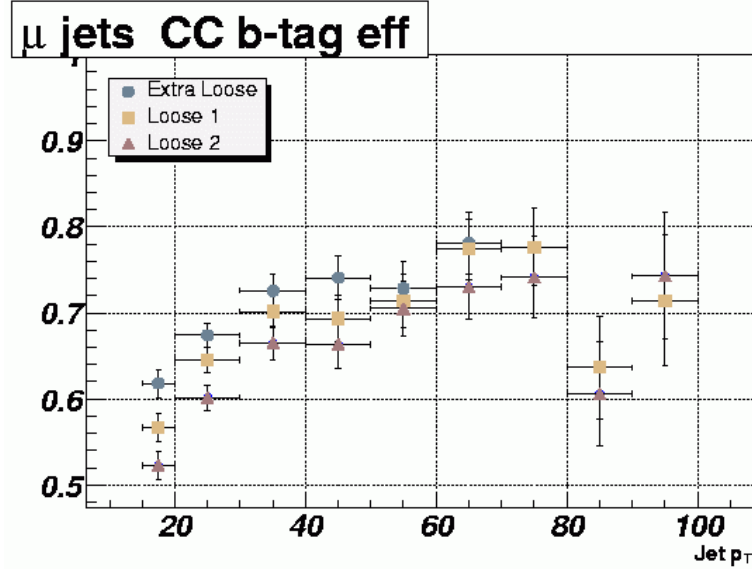
Figure 4.22: The b–tagging efficiency in data (after taggability requirements), as a function of jet E_T, for central jets ($|\eta|<1.2$). This analysis uses the Extra Loose secondary vertex definition, which has the highest b–jet tagging efficiency.

Carlo samples (Pythia inclusive QCD).

Figure 4.22 shows the performance of b–tagging in data for taggable jets in the central region ($|\eta|<1.2$) as a function of jet E_T.

4.5.4   Scale Factors

In simulated events, where the true particle content of jets is known, measuring the efficiency and fake rate of the b–tagging algorithm is trivial. The efficiency is the number of b–jets which are b–tagged divided by the number of taggable b–jets. A reconstructed jet in MC is defined as a b–jet when a b–hadron exists within $\Delta R<0.3$ of the jet [15]. The fake rate is the number of light jets which are b–tagged divided by the number of taggable light jets. A reconstructed jet in MC is defined as a light jet

---

[15]Greater than 99% of b–jets in simulated Z$\to b\bar{b}$ events have b–hadrons within $\Delta R<0.3$ of the jet axis.
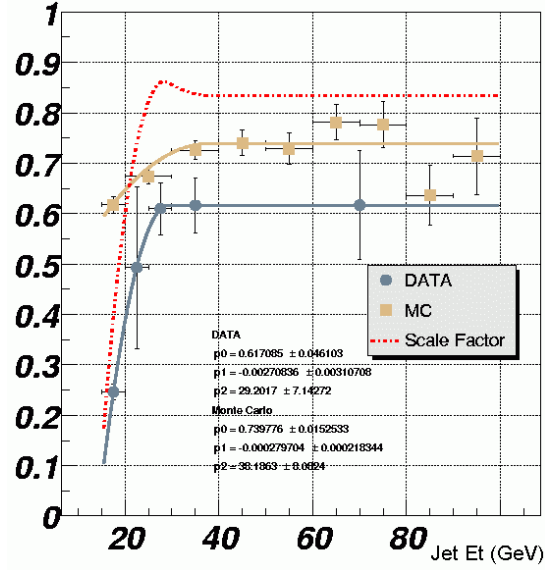
Figure 4.23: The b–tagging efficiencies per taggable b–jet in data and MC events, and the scale factor derived between them, as functions of jet E_T.

if no b– or c–hadrons exist anywhere in the event.

The performance of the b–taggers in the detector simulation is not the same as in the data, so a scale factors is applied to b–jets in the Monte Carlo to better model the real behavior in data. The average scale factors are 0.78±0.06 in the central region ($|\eta|$<1.2) and 0.63±0.14 in the forward region ($|\eta|$>1.5). The scale factor is shown as a function of b–jet E_T in the central region in Figure 4.23. Uncertainties in the scale factor take into account all statistical errors from the finite number of events in the MC and μ–in–jet data samples and all systematic errors from the methods used to derive the b–tagging efficiency in MC and data.

Figure 4.24 shows the ratio of c–tagging efficiency to b–tagging efficiency, derived in MC simulation, for taggable jets in the central region. The c–tagging efficiency will be used to estimate the size of a $c\bar{c}$ background in Chapter 7.
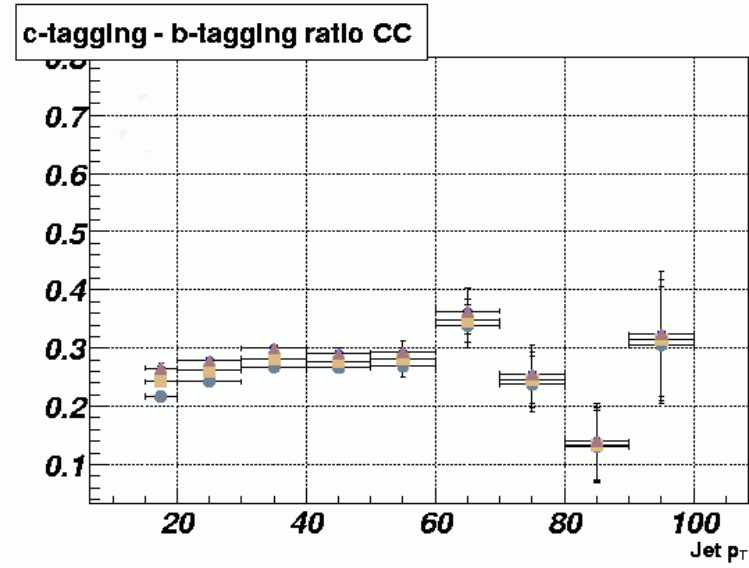
Figure 4.24: The ratio of the c–tagging to the b–tagging efficiency in data, as a function of jet E_T.

4.5.5 Multi–jet Biases

Monte Carlo events are studied to see if the b–tagging efficiency for a b–jet opposite to a jet that is already b–tagged is different than for an unbiased b–jet, correcting for the jet E_T and $|\eta|$. The b–tagging efficiency for such jets is found to be higher by a factor of 1.03±0.01(stat), within the assigned systematic uncertainty for b–tagging efficiency.

The b–tagging efficiency and fake rate may change as a function of the number of jets in the event. Measurements of the b–tagging efficiency and fake rate in data, and their comparisons to Monte Carlo, are performed in di–jet samples. The scale factor between data and Monte Carlo is verified to not be sensitive to the number of jets in the event. As shown in Figure 4.25, the negative tagging rate in the data sample (which is proportional to the rate of fake b–tags) is stable as a function of the number of jets in the event. Plotting the negative tag rate as a function of the p_T^{rel} of the muon in the jet shows that there is no correlation with the fraction of
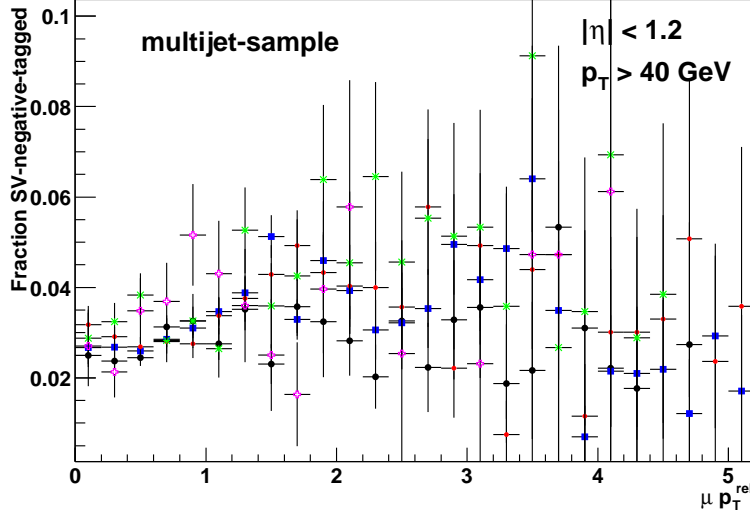
Figure 4.25: The negative b–tagging efficiency of a jet as a function of the p_T^{rel} of the muon in the jet, for various jet–multiplicity events, for n_j of 2 (black), 3 (blue), 4 (red), 5 (green) and 6 (pink).

true b–jets in the events, since the fraction of true b–jets peaks at a p_T^{rel} of about 1.5 GeV/c. This suggests that no modification to the scale factor needs to be applied as the jet multiplicity increases, at least for events with no more than six jets [16]. This conclusion is verified by the MC simulation of the b–tagging efficiency per jet for events with different numbers of jets, as shown for a simulated neutral Higgs boson signal in Figure 5.14.

---

[16]This analysis does not use events which have more than five reconstructed jets.

Chapter 5

SIMULATION

Samples of expected backgrounds and hypothesized signals are produced using Monte Carlo programs and a full detector simulation. The simulated signals are used to determine the efficiency for observing the hypothesized neutral Higgs boson production. The simulated backgrounds are used to check the understanding of the detector and reconstruction performance, by comparing observed events to expectations from the SM.

5.1 Monte Carlo

Events of the expected signals are generated with Pythia [19] (version 6.202), which has two parts. The first allows the user to select one of the hard–coded matrix elements to calculate hard–scattering interactions of the type desired. The second half performs the *showering* of final–state colored particles into jets [1] and the decay of short–lived particles (such as b–hadrons). Multi–jet background event samples use the ALPGEN [20] (version 1.2) generator to calculate final–state colored particles, which are then passed through Pythia showering. The ALPGEN generator calculates many SM processes at hadron colliders exactly, including multi–jet production, at leading–order. ALPGEN is able to calculate complicated interactions with many (up to 10) jets in the final state. The results of ALPGEN are also compared to cross–sections calculated using MADGRAPH [21], another exact leading–order generator. MAD-GRAPH calculates, by "brute force", the amplitudes for all leading–order processes

---

[1]Pythia also performs a "backwards showering" to simulate initial state radiation (ISR) from the incoming colored particles.

consistent with the SM which have the specified initial and final states.

Events generated with Pythia use the CTEQ4L parton distribution functions (PDF), and events generated using ALPGEN or MADGRAPH use CTEQ5L. Total cross–sections for the generated processes are always taken from published next–to–leading–order calculations, where available, or normalized to data.

After Pythia showering, events are created by a full DØ detector simulation. Software based on the GEANT software package (version 3) simulates the propagation and material interactions of the particles as they pass through the detector. Minimum–bias events (random $p\bar{p}$ inelastic interactions), generated with Pythia are added to the hard–scatter events. The number of minimum–bias events added is Poisson distributed with a mean of 0.5, corresponding to the approximate instantaneous luminosity at which the data was taken $(1 - 4 \times 10^{31}$ cm^{-2}s$^{-1})$. The simulated ionization in the detector is passed into a simulation of the detector's electronic responses and digitization. The resulting events are identical in form to those read out from the real detector. The same detector reconstruction algorithms used for the real data are then applied to the simulated events. Finally, the ROOT software package produces files for analysis.

5.1.1 Signals

There are two related leading–order processes for neutral Higgs production which give a multi–jet signal with three or more high–p_T b–jets. The first is $b\bar{b}$h$\rightarrow b\bar{b}b\bar{b}$, where four b–quarks exist at high p_T (defined as $p_T>15$ GeV/c). The other is the leading–order process bg$\rightarrow$bh$\rightarrow$b$b\bar{b}$, where only three b–quarks exist at high p_T. After demanding at least three b–tags, the two signals have very similar acceptances, as seen in Figure 5.1, which shows the invariant mass distribution of the leading two jets (m_{01}) for the bh and bbh signals, after requiring three b–tagged jets.

If two processes create physically different final states, then they are independent and can simply be combined. If the final states created by the two processes are
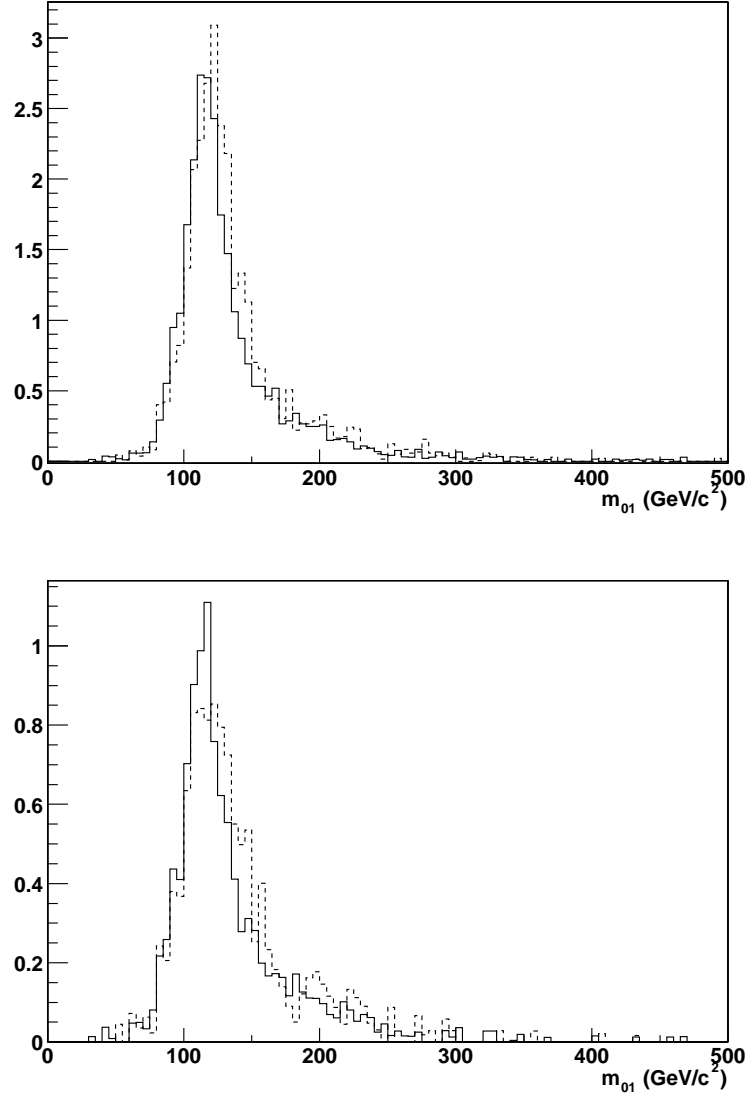
Figure 5.1: Comparison of the invariant mass distribution of the leading two jets (m_{01}) for the bh (solid) and bbh (dashed) signal processes, after requiring 3 b–tagged jets. The top plot requires only ≥ 3 jets, and the bottom plot requires ≥ 3 jets.

indistinguishable for some fraction of the events, then combining the event samples would be *double-counting.* At LO, with initial-state radiation (ISR) and final-state radiation (FSR) turned off in Pythia, the bh and bbh *spectator b–quarks* (the b–quarks from the initial state which do not radiate a Higgs boson) have very different properties, as seen in Figure 5.2. The bh process always has a spectator b–quark in this case at very high $|\eta|$, whereas all b–quarks in the bbh process are more central. Thus, it would seem that the signals could be combined without double-counting. However, when ISR and FSR are included, the signals are no longer significantly different, as seen in Figure 5.3.

Because the two leading–order signal processes, bh and bbh, overlap kinematically after radiation effects, a procedure must be chosen to avoid double counting [13] [14]. One solution is to match the two signals by the p_T of the lowest p_T spectator b–quark. Events from the bh sample with the lowest p_T spectator b–quark below a given p_T cut are kept, and those from the bbh sample with a lowest p_T spectator b–quark above the given p_T cut are kept. All other events are discarded. This solution takes advantage of the idea that the bh signal more accurately models the kinematics when the p_T of the lowest p_T spectator b-quark is low, whereas the bbh signal is a better model for signal kinematics when it is high [15]. The effects on the signal distributions from following this prescription are shown in Figure 5.4, using a p_T for matching of 15 GeV/c.

The other solution is simply to use only one of the LO signal processes. This certainly avoids double counting, and is the most conservative approach. Because the signals are nearly identical, the signals are each very nearly cut in half by the p_T matching prescription, as seen in Figure 5.4. Very little signal ($<5\%$) is lost by using the "one signal only" prescription. The bh signal is a simpler calculation at NLO theoretically, and suffers from smaller scale uncertainties at LO. Thus, in this analysis, only the bh signal is used for the final neutral Higgs boson search. The bbh signal samples are still useful however, for performing various studies and cross–checks.
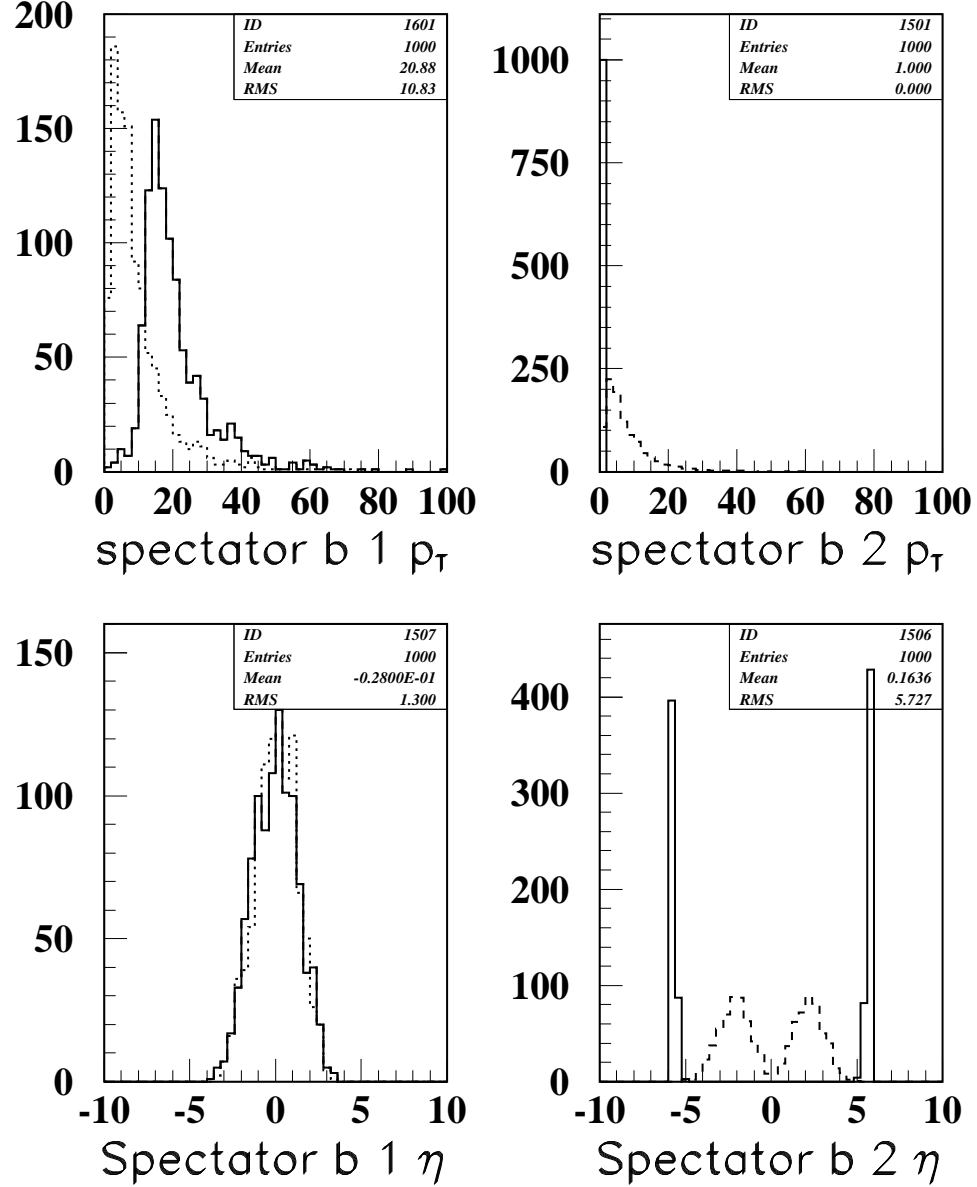
Figure 5.2: A comparison of the p_T and η spectra of the bh (solid) and bbh (dashed) MC signals from Pythia with no ISR or FSR. The spectator b 1 is the b parton which radiated the Higgs, and is thus at higher p_T and more central, typically than spectator b 2 which is simply from the initial "gluon splitting" and does not interact further. Without ISR and FSR, the spectator 2 b parton has very different kinematics in bh (where it basically goes down the beam-pipe) and bbh.
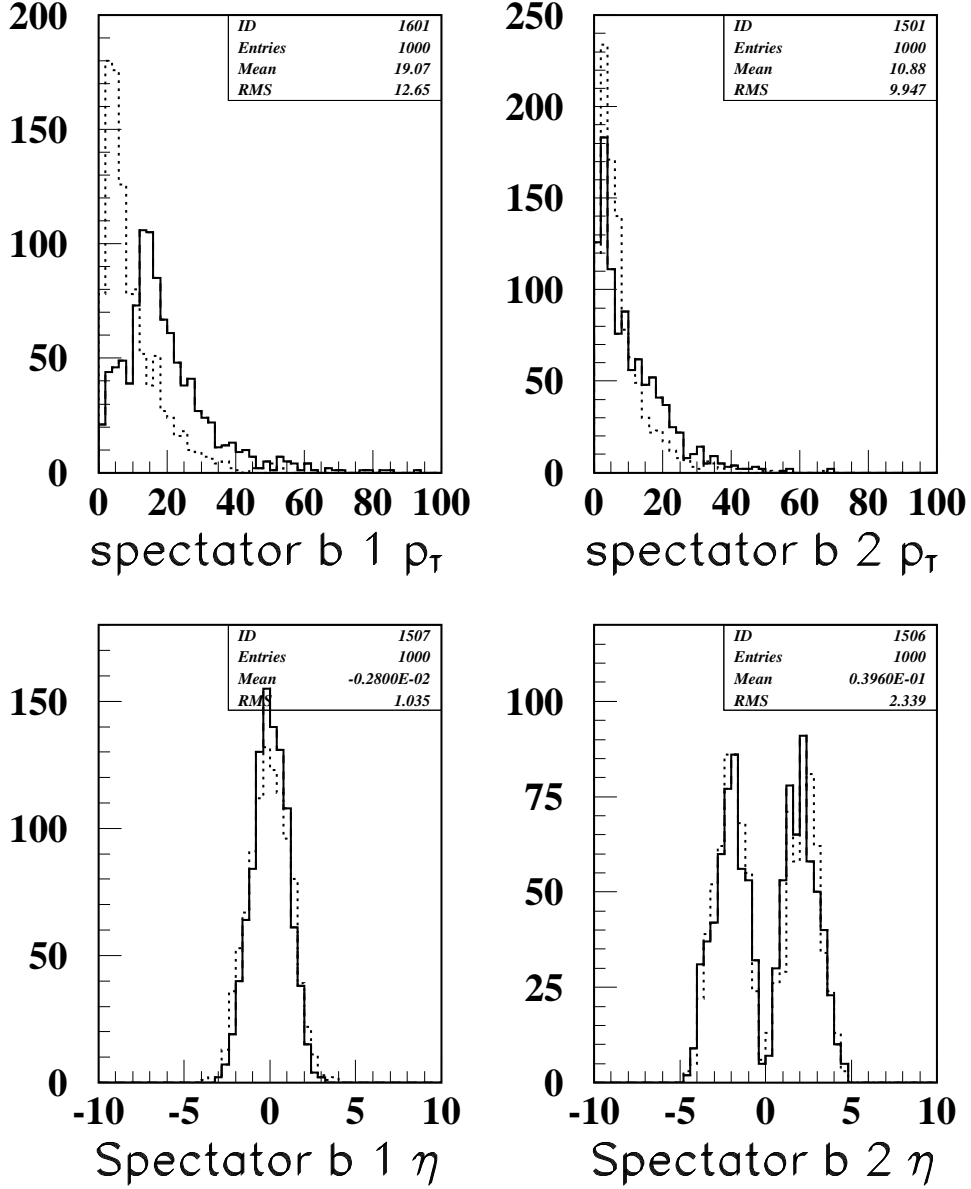
Figure 5.3: A comparison of the p_T and η spectra of the bh (solid) and bbh (dashed) MC signals from Pythia with ISR and FSR turned on. The spectator b 1 is the b parton which radiated the Higgs, and is thus at higher p_T and more central, typically than spectator b 2 which is simply from the initial "gluon splitting" and does not interact further, aside from radiation. With ISR and FSR, the spectator 2 b parton has very similar kinematics in bh and bbh.
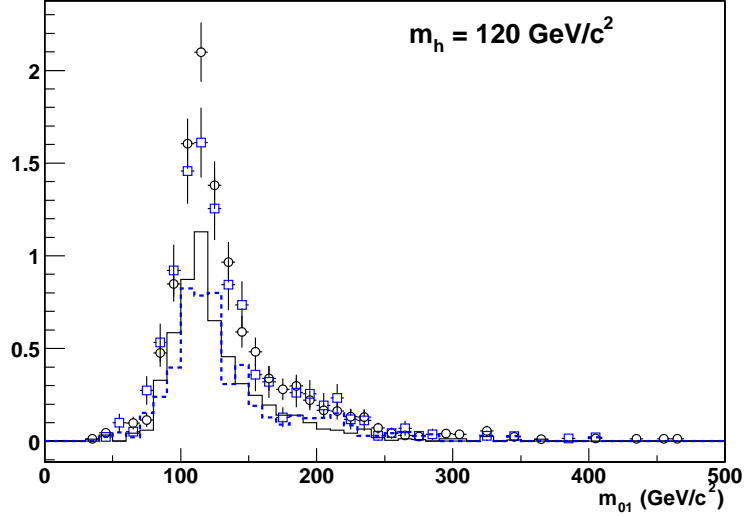
Figure 5.4: The invariant mass of the two leading jets for the bh (circles and solid) and bbh (squares and dashed) signals with 3 b–tags in the 4–jet channel with $m_h = 120$ GeV/c^2 before (points) and after (histograms) applying the p_T matching prescription.

bh Samples

The process gb→bh→b$b\bar{b}$ is implemented as process 32 in Pythia. The Higgs boson in each event is required to have $p_T>15$ GeV/c, but no rapidity cuts are used. Samples of 25,000 events are made for each Higgs mass of 100, 120, and 150 GeV/c^2. The p_T and rapidity spectra of the Higgs (for $m_h = 120$ GeV/c^2) are compared to those from the NLO calculation [15] using the authors' program, MCFM. The shapes show decent agreement, as seen in Figures 5.5 and 5.6. This indicates that the Pythia leading–order kinematics are roughly correct, especially after ISR and FSR effects in Pythia, which mimic the behavior of higher–order processes (such as gluon radiation). However, the shapes will be further matched to the NLO calculation, as described below. It was verified that the p_T spectrum of Pythia without ISR had a hard cut at the generated hard–scatter p_T of 15 GeV/c.

The total NLO values for the signal cross–sections are used (see Table 5.1). Each
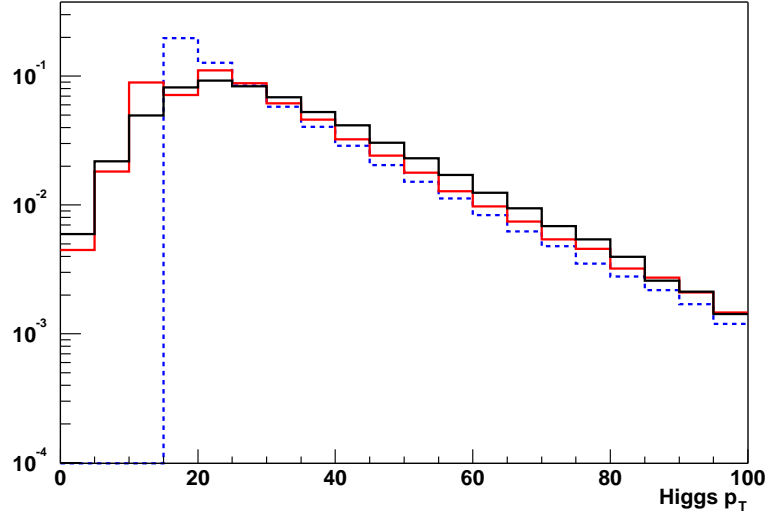
Figure 5.5: Comparison between the p_T spectra of the simulated Higgs in MCFM at LO (dashed–blue), MCFM at NLO (solid–red), and our Pythia simulation (solid–black).
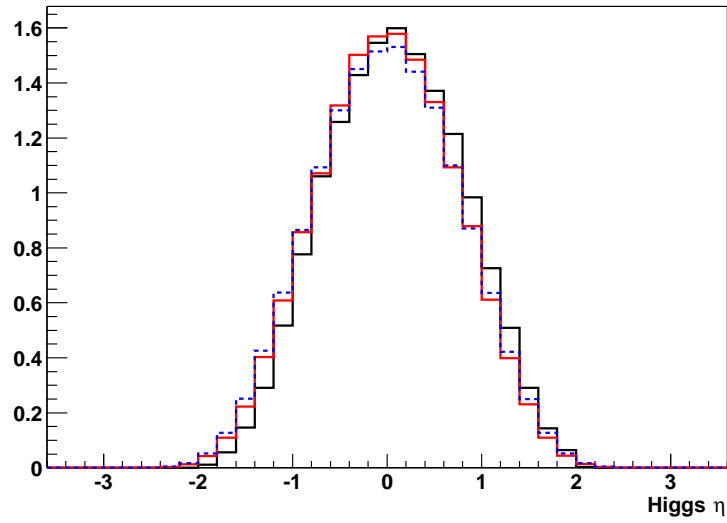
Figure 5.6: Comparison between the η spectra of the simulated Higgs in MCFM at LO (dotted–blue), MCFM at NLO (solid–red), and our Pythia simulation (solid–black).
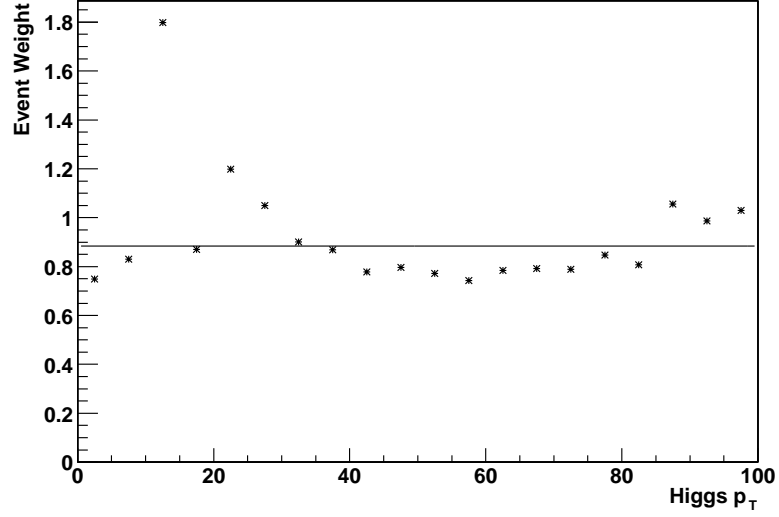
Figure 5.7: The weight given to each event in the bh Monte Carlo such that the correct NLO Higgs p_T spectrum is reproduced. Also shown is a fit to a constant (the average weight).

leading–order event is weighted, based on the p_T of the Higgs in the event, such that the resulting p_T spectrum for the simulated Pythia leading–order events exactly reproduces the NLO spectrum (from MCFM). The weighting function used is shown in Figure 5.7. The effect on the total number of events in the signal peak is shown in Figure 5.8. There is a 13% reduction in the total number of events passing analysis cuts caused by the re–weighting of the events to match the NLO Higgs p_T spectrum. The uncertainty due to the re–weighting is considered along with all other signal generation errors when determining systematic errors (see Chapter 7).

bbh Samples

The process $b\bar{b}h{\rightarrow}b\bar{b}b\bar{b}$ is process 121 in Pythia. 25,000 events are generated for each Higgs mass, 100, 120, and 150 GeV/c^2. Very recently, NLO calculations have been performed [16][17]. There has been great success in understanding the factorization
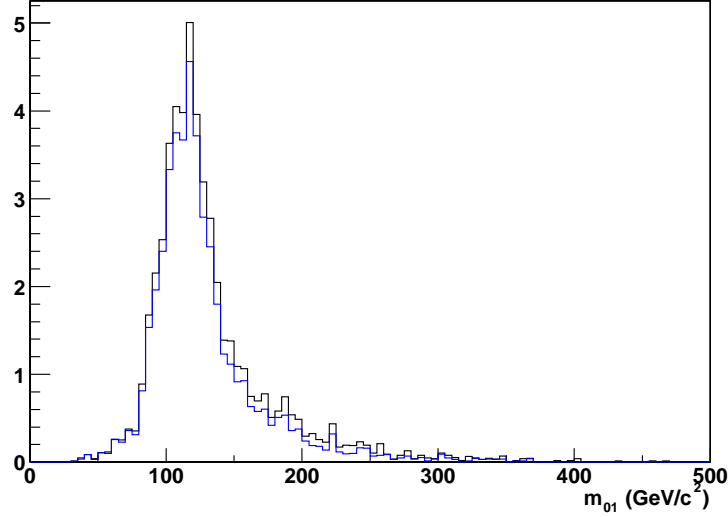
Figure 5.8: Comparison between the number of events in the non–b–tagged invariant mass peak of the bh Monte Carlo sample (with m_h=120 GeV) passing cuts before (black, higher) and after (blue, lower) re–weighting the leading–order events to match the MCFM NLO Higgs p_T spectrum.

and renormalization scale appropriate for generating bbh at LO. The results from LO studies are now corroborated by the NLO result. Previous theoretical uncertainties of a factor of $\gtrsim 5$ in cross–section are now reduced to 25% [18]. Events are generated with no cuts on the p_T's or rapidity of the bottom quarks. Cross–sections are taken from the NLO calculations, and are listed in Table 5.1.

Signal Simulation Quality Checks

The di–jet invariant masses of the leading two E_T jets, jet kinematics, and b–tagging results are studied in the signal simulation. The results are as anticipated for neutral Higgs boson production. Plots showing the same basic quantities checked for data quality are shown in Figures 5.9 and 5.10 for a bbh Monte Carlo sample (with $m_h = 120 \, \text{GeV}/c^2$). Figure 5.11 shows the di–jet invariant mass spectrum of the leading two E_T jets for each of the generated Higgs masses. The b–tagging performance in the

Table 5.1: The cross–sections for signal production, at $\tan\beta$ of 1, used to normalize the Monte Carlo samples. The values are taken from the NLO calculations. (See the text for details.)

m_h(GeV/c^2)	bh (fb) (p_T>15 GeV/c)	bbh (fb) (no cuts)
100	6.45	19.8
120	3.03	8.87
150	1.14	3.04

signal Monte Carlo is shown in Figures 5.12, 5.13, and 5.14 for a bbh sample (with m_h = 120 GeV/c^2) as a function of the jet E_T, η, and number of jets (n_j), respectively.
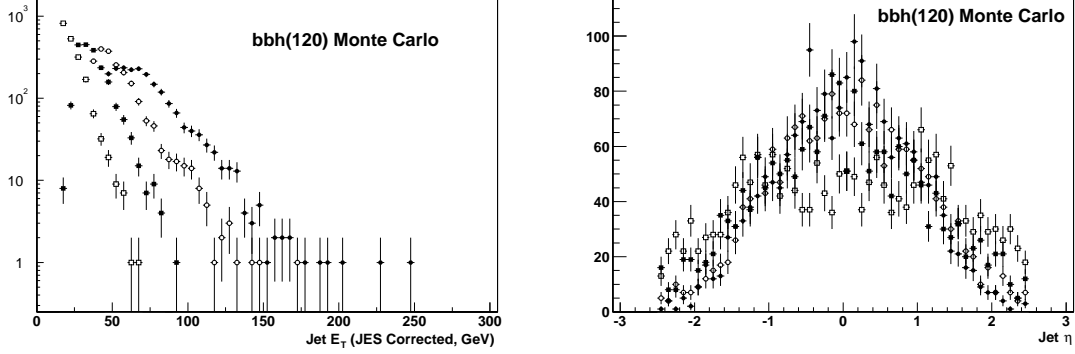
Figure 5.9: Left: The E_T distributions of the first (solid circles), second (open circles), third (solid squares), and fourth (open squares) highest E_T jets in each event of the generated bbh Monte Carlo sample with $m_h = 120\,\text{GeV/c}^2$. Right: The η distributions of the first (solid circles), second (open circles), third (solid squares), and fourth (open squares) highest E_T jets in each event of the generated bbh Monte Carlo sample with $m_h = 120$ GeV/c^2. For both plots, jets must pass all quality cuts and taggability requirements, each event must pass one of the (simulated) multi–jet triggers, jet E_T's are corrected for jet–energy scale, and the events must pass E_T cuts of 45, 35, and 15 GeV for the first, second, and third leading E_T jets.

Figure 5.10: Left: The distribution of the number of jets in each event of the generated bbh Monte Carlo sample with $m_h = 120\,\text{GeV/c}^2$. Right: The $\eta - \phi$ of each jet in each event of the generated bbh Monte Carlo sample with $m_h = 120$ GeV/c^2. For both plots, jets must pass all quality cuts and taggability requirements, each event must pass one of the (simulated) multi–jet triggers, jet E_T's are corrected for jet–energy scale, and the events must pass E_T cuts of 45, 35, and 15 GeV for the first, second, and third leading E_T jets.
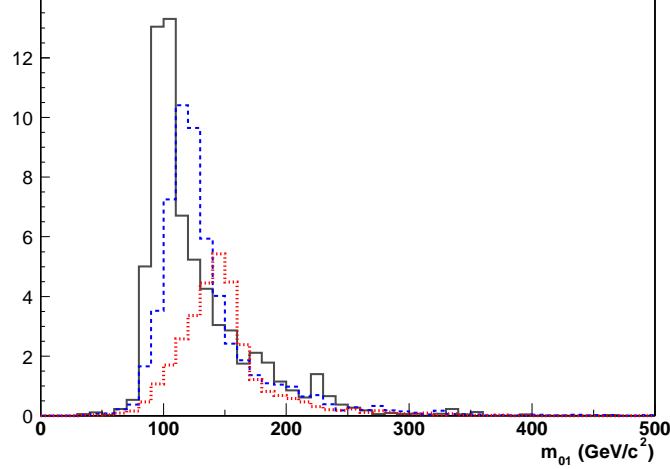
Figure 5.11: The invariant mass spectrum formed from the leading two E_T jets (JES corrected) for each of the Higgs masses generated ($m_h = 100$, 120, and 150 GeV/c^2). A $\tan\beta$ of 50 is assumed, for calculating the cross–sections.

Signal Di-jet Mass Shapes

The characteristics of the reconstructed mass of the Higgs are studied in Monte Carlo events using both the bh and bbh signal samples for Higgs masses of 100, 120, and 150 GeV/c^2. The invariant mass of the leading two E_T jets is calculated for events which have three b–tags, as shown in Figure 5.15. Most events with a reconstructed invariant mass in the high tail above the Gaussian peak are due to events where the largest two reconstructed E_T jets are not from the decay of the Higgs boson. Between 80–90% of events have the correct assignment of jets from the Higgs as the leading two E_T jets, depending on m_h. The higher the mass of the Higgs boson, the more likely the leading E_T jets are to be from its decay. The low tail of the Gaussian shape of the reconstructed invariant mass spectrum is mostly due to final-state radiation (FSR) of jets which came from the decay of the Higgs boson.

Other combinations of jets, such as the second and third leading E_T jets or an average of the invariant masses between all three combinations of the three leading
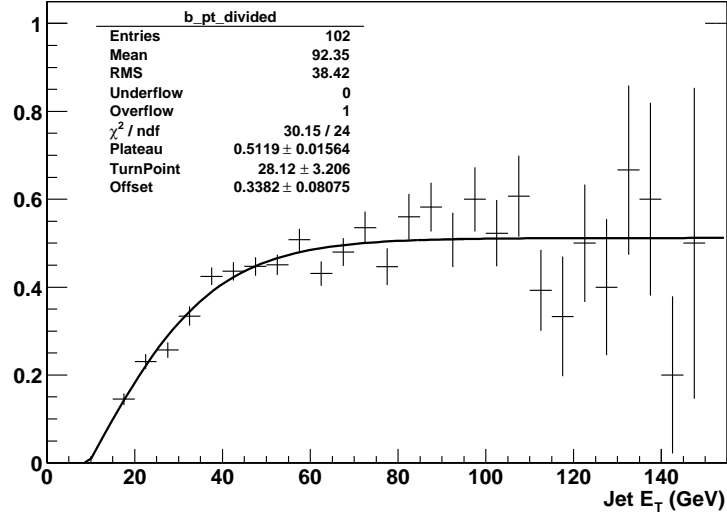
Figure 5.12: The b–tagging efficiency for a jet, as a function of the E_T of the jet (JES corrected), in the bbh Monte Carlo sample (with $m_h = 120$ GeV/c^2).
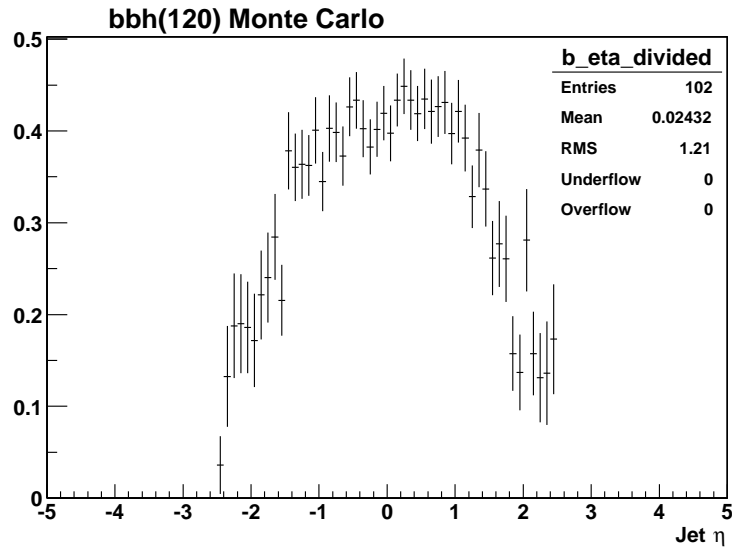
Figure 5.13: The b–tagging efficiency for a jet, as a function of the η of the jet, in the bbh Monte Carlo sample (with $m_h = 120$ GeV/c^2).
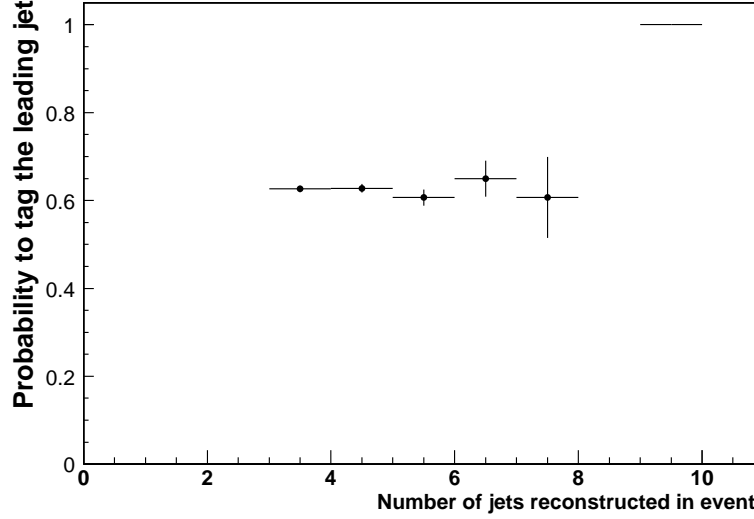
Figure 5.14: The b–tagging efficiency for the highest E_T jet in each event (JES corrected), as a function of the number of jets reconstructed in the event, in the bbh Monte Carlo sample (with $m_h = 120$ GeV/c^2).

E_T jets, are studied and found to give wider signal peaks. An example is shown in Figure 5.16, where the average invariant mass of the three combinations of the three leading E_T jets is shown.

5.1.2 Backgrounds

Any process which creates events with high jet multiplicity, particularly heavy–flavor (c– or b–quark) jets, is a background to a triple b–tagged final state with three or four jets. Chapter 7 describes in detail the methods used for normalizing the difficult multi–jet backgrounds using data. This Section lists the processes studied and explains the tools used to create the simulated events.
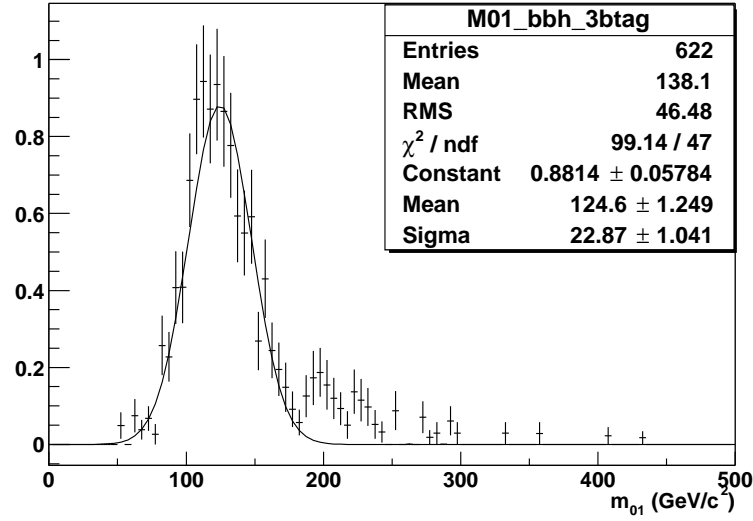
Figure 5.15: A Gaussian fit to the invariant mass of the two leading E_T jets in the bbh Monte Carlo sample, with $m_h = 120$ GeV/c^2.
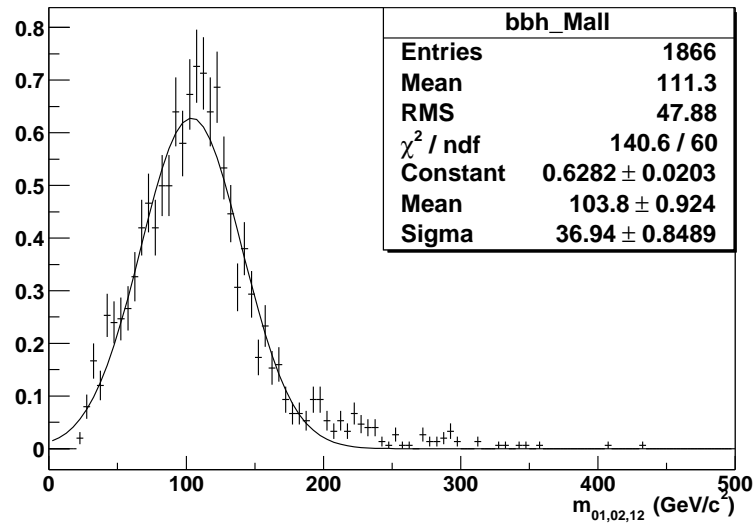
Figure 5.16: A Gaussian fit to the average invariant mass of the three combinations of the three leading E_T jets in the bbh Monte Carlo sample, with $m_h = 120$ GeV/c^2.

Table 5.2: Background Monte Carlo cross–sections.

Process	Cross–section (pb)	Generator Cuts (p_T in GeV/c)
$b\bar{b}$jj (ALPGEN)	1568	p_T(b)>30, p_T(j)>15, $|\eta|$<3.0, ΔR>0.4
$b\bar{b}$jj (MADGRAPH)	2655	p_T(b)>30, p_T(j)>15, $|\eta|$<3.0, ΔR>0.4
$b\bar{b}$j (ALPGEN)	3842	p_T(b)>30, p_T(j)>15, $|\eta|$<3.0, ΔR>0.4
$b\bar{b}b\bar{b}$ (ALPGEN)	51	p_T(b)>15, $|\eta|$<3.0, ΔR>0.4
$t\bar{t}$ (Pythia)	7	none
Z($b\bar{b}$)+jets (Pythia)	1180	none
Zb (MADGRAPH)	10	p_T(b)>15, $|\eta|$<3.0, ΔR>0.4
Z$b\bar{b}$ (MADGRAPH)	3	p_T(b)>15, $|\eta|$<3.0, ΔR>0.4
Zb(b)+mistakes (NLO)	40	p_T(b)>15, $|\eta|$<2.0, ΔR>0.7

Heavy–Flavor Multi–Jet

Multi–jet production, such as $p\bar{p}\rightarrow$3j, 4j, $b\bar{b}$j, $b\bar{b}$jj, and $b\bar{b}b\bar{b}$ (where j represents a light quark (u,d,s), charm quark, or gluon jet) is very difficult to model correctly, due to the large number of diagrams contributing to the complicated final state. Processes with heavy–flavor jets in the final state have the added difficulty of correctly modeling the observed rates and kinematic distributions of the heavy–flavor jets. Fortunately, full leading–order matrix element generators, are now able to produce events of these complicated processes, using relatively little computer time. Using the exact leading–order matrix elements is more accurate than modeled gluon radiation obtained with Pythia $p\bar{p}\rightarrow b\bar{b}$ + jets production, for example.

ALPGEN is used to generate 50,000 events each of $b\bar{b}$j, $b\bar{b}$jj, and $b\bar{b}b\bar{b}$. Generator-level cuts are applied to the outgoing partons in order to leave sufficient statistics after analysis cuts. The outgoing b–quarks are required to have p_T>30 GeV/c and the light jets p_T>15 GeV/c, all with $|\eta|$<3, except for the $b\bar{b}b\bar{b}$ samples, which required all b–quarks to have only p_T>15 GeV/c. A ΔR>0.4 requirement is made on all outgoing partons, to reflect the 0.5 cone size of our jet reconstruction algorithm. The cross–

sections given by ALPGEN for these processes are listed in Table 5.2. In Chapter 7, comparisons with data are made, and normalization for these processes are derived.

A bug in the ALPGEN to Pythia interface was recently uncovered, in which no underlying event was generated. Small samples of $b\bar{b}$j events without this bug were produced, and there were negligible acceptance differences between the samples with and without the bug. Making new samples without the bug was determined not to be worth the large amount of computer time necessary.

The $b\bar{b}$jj cross–section is also checked with another leading–order matrix–element Monte Carlo generator, MADGRAPH, using identical generator–level cuts as for ALPGEN. There are over 5,000 leading–order processes calculated for $b\bar{b}$jj. The result is also shown in Table 5.2. The extent to which the two calculations for the $b\bar{b}$jj cross–sections agree gives a feel for the error associated with these calculations. They can be trusted to within factors of $\sim$2, but should not be expected to model data more accurately. These expectations are confirmed in our comparison of the ALPGEN Monte Carlo to multi–jet data in Chapter 7.

Other Backgrounds

Other processes are also considered as possible backgrounds, and are simulated. 50,000 inclusive $p\bar{p}{\rightarrow}t\bar{t}$ events are generated with Pythia. A production cross–section of 7 pb is assumed (as extrapolated in $p\bar{p}$ center–of-mass energy from Run I and verified by preliminary results from Run II). There are always two b–quarks from the decays of the top quarks, and often two to four other jets, if one or both of the W bosons from the top quarks decays hadronically. Jets from the W bosons can be b–jets and c–jets, easily misidentified as b–jets. Even if the W decays to light–quark jets, these can be misidentified as b–jets a small fraction of the time. Thus all $t\bar{t}$ final decay states are simulated.

20,000 events are made using Pythia of $p\bar{p}{\rightarrow}Z({\rightarrow}b\bar{b})$ +jets, i.e. the Z is forced to decay to $b\bar{b}$. The cross–section used for the process is 1.18 nb, as measured by DØ

in Run II through Z→$\mu^+\mu^-$ [22], accounting for the ratio of the Z→$b\bar{b}$ to Z→$\mu^+\mu^-$ branching fractions in the SM. The extra jets are generated by Pythia, which most likely do not accurately model the true radiated E_T spectra (Pythia radiation in such cases is usually too soft). However, it is fair to use the sample to estimate the shape and position of the Z peak, since this does not heavily depend on the radiation. The events are also mainly used as a cross–check for the shape of the double b–tagged data, where the events with extra jets which have been b–tagged make up a small contribution, since the b–tagging efficiency for real b–jets is much larger than for light–jets.

30,000 events of Zb (with Z→$b\bar{b}$) were generated with Pythia [2]. The cross–section for Zb production has recently been calculated at NLO [23], so this value is used, listed in Table 5.2. The cross–section includes contributions from Zb (13pb), Zbb (6pb), Zc and Zcc fake b–tags, and Z + light–jet fake–tags (assuming a fake b–tagging rate of 0.5% for light–quark (u,d,s) and gluon jets and 15% for c–jets). Some of the Z processes' cross–sections are also calculated with the MADGRAPH generator, and are included in Table 5.2.

---

[2]Zb is made by using MSEL=13 in Pythia. The KFIN parameters of Pythia are used to set all incoming u,d,s,c quark distributions to zero, thus selecting only gb→Zb.

Chapter 6

TRIGGER AND DATA

6.1 Trigger

The total cross–section for multi–jet events is very large compared to the expected neutral Higgs boson signal. A specialized set of trigger requirements is used to accept the largest possible fraction of signal events, while remaining within rate–to–tape constrains ($<$4 Hz), up to high instantaneous luminosity (4×10^{31} cm^{-2}s^{-1}). The data collected spans two trigger versions (v10 and v11), with the first collecting about 75 pb^{-1} of integrated luminosity and the second about 56 pb^{-1}. [1]

The first trigger version required four calorimeter towers with E_T>5 GeV at L1, three L2 jets with E_T>8 GeV and total L2 H_T (scalar E_T sum of L2 jets with E_T>5 GeV) above 50 GeV, and three L3 jets with E_T>15 GeV. The η_d coverage was up to 2.4 at L1/L2 and 3.0 at L3.

The second trigger version required only three calorimeter towers with E_T>5 GeV at L1, the same requirements as the v10 trigger at L2, and three L3 jets with E_T>15 GeV, where two of them have E_T>25 GeV.

The L3 jets used in both triggers were 0.5 cone jets, not corrected for non–linearity of the calorimeter, or jet energy scale. However, in the second trigger list, the jets' E_T and η were corrected for the Z position of the primary vertex, significantly sharpening their E_T and η resolution.

Accurately calculating the efficiency of each complete trigger for the expected

---

[1]The sum of the integrated luminosity, 131 pb^{-1}, was less than the 141 pb^{-1} recorded to tape on all triggers, due to occasional limitations imposed at the highest instantaneous luminosities.

signal and backgrounds is essential. Significant effort is devoted to using data to understand the performance of the triggers, since even detailed software simulations of the trigger systems (TrigSim) may be inaccurate in such complicated multi–jet events. [2] The probability for each event to pass each trigger requirement is calculated, as a function of the properties of the objects reconstructed offline, such as the offline jets' E_T's. These individual probabilities are then combined into total probabilities for each simulated event to pass all trigger requirements. This total trigger probability is applied as a weight to each simulated event, as described below.

6.1.1   L1 and L2 Triggers

Detailed studies of the L1 and L2 triggers in multi–jet events, and derivations of the methods used to calculate the trigger performances at L1 and L2, are documented in Appendix A.

The L1 trigger terms used for this analysis are based on calorimeter trigger towers. For each event, a probability is calculated that enough towers would be above the trigger thresholds, based on the E_T's of the offline reconstructed jets. A study is made of jets and L1 trigger information in data recorded with a trigger requiring only a single muon. The muon triggered data is as unbiased as possible towards jets [3]. The number of L1 towers above threshold near offline jets is measured, as a function of offline jet E_T. This distribution, shown in Figure 6.1, is then used to calculate the total probability for a given number of towers to be above threshold for a multi–jet event. A closure test is performed which verifies that the predicted probability for passing the trigger requirements agrees with the actual pass fraction of the online trigger. The data used for the closure test is independent from the data used to

---

[2]TrigSim is compared with the trigger information from data, and used to cross–check the methods and assumptions of the efficiency studies. The comparisons shows good agreement.

[3]Minimum–bias triggered events are completely unbiased (by definition), but not a large enough sample of events is recorded for performing detailed trigger studies.
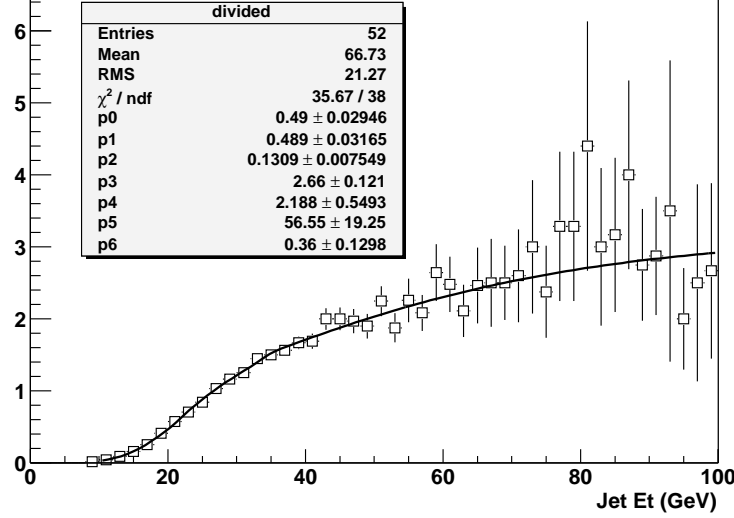
Figure 6.1: The average number of calorimeter towers above 5 GeV associated with a jet as a function of the E_T (uncorrected) of the jet.

derive the probability distributions. As seen in Figure 6.2, the predicted total trigger pass probabilities match the actual pass fractions of the trigger well. Results are also compared with those from TrigSim, which shows good agreement.

The L2 trigger requirements used in this analysis are based on clusters of calorimeter trigger towers, called L2 jets. Using the same data set used to study the L1 trigger terms, but requiring the L1 trigger requirements to have already been satisfied, turn–on curves are measured for the L2 jets as a function of offline reconstructed uncorrected jet E_T. The distribution of the number of "noise jets" (those not matched to any offline jets) is also measured. Using these curves and the noise jet distribution, the total probability for passing a trigger requiring a given number of jets to be above a certain threshold is calculated. As in the L1 study, a closure test is performed which compares the calculated probability for a given number of L2 jets to be above threshold with the actual pass fraction of the online L2 trigger, as seen in Figure 6.3. The results are also compared to the TrigSim calculations.
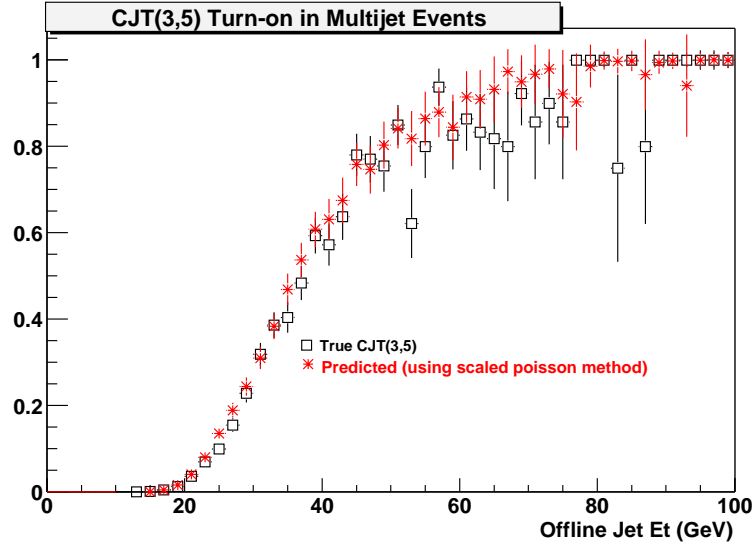
Figure 6.2: Comparison between the predicted probability of satisfying the CJT (3,5) L1 trigger and the actual probability, measured on the independent muon–triggered data sample, as a function of the E_T (uncorrected) of the highest E_T (uncorrected) offline jet.

6.1.2 L3 Trigger

The method used for calculating the total efficiency of the L3 trigger is identical to that used for calculating the total efficiency for passing the L2 jets requirement. Turn–on curves are measured in data for L3 jets, as a function of offline reconstructed un-corrected jet E_T. Although the same data is used to reconstruct jets at L3 and offline (the zero–suppressed, full–readout, calorimeter cell data), algorithms at L3 are simpler and less precise, to save processing time. Most notably, there is no splitting or merging of jets in L3, or corrections for noisy cells and non–linearity, as opposed to the offline jet reconstruction (see Chapter 4). Also, for the first trigger version (v10), the primary vertex Z position was assumed to be the center of the detector, thus smearing the actual E_T of the L3 jets.

The turn–on curves for L3 jets, such as the one shown in Figure 6.4, are measured
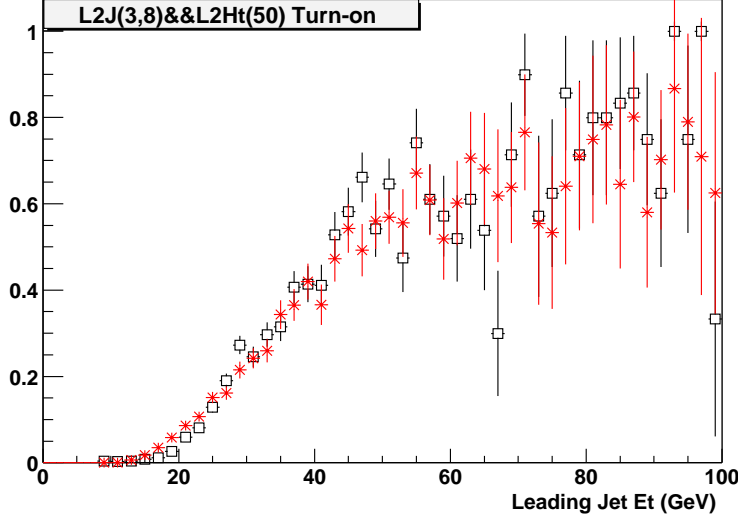
Figure 6.3: Comparison of the predicted probability of satisfying the L2(3,8) && L2Ht(50) L2 trigger and the actual probability, measured on the independent muon–triggered data sample, as a function of the E_T (uncorrected) of the highest E_T (uncorrected) offline jet.

as the fraction of offline jets which have a L3 jet with $E_T > 15$ GeV matched to them within $\Delta R < 1.0$. The curves are parameterized with the following form of the tanh function:

$$C(x) = P_0 + P_1 tanh((x - P_2)/P_3)$$

Using these turn–on curves, the total event probability of having three jets above threshold can be calculated, by combining the individual jet probabilities, assuming the jets are independent. The validity of this assumption is demonstrated by comparing the predicted pass fraction for multi–jet events with the actual pass fraction, in an independent data set, as a function of the leading offline jet E_T (as shown in Figure 6.5) or the total offline jet $\mathcal{H}_T$.
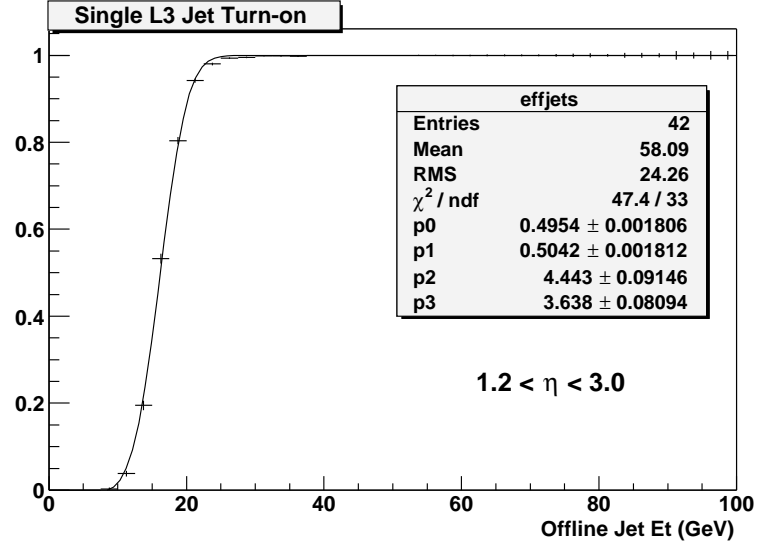
Figure 6.4: Turn–on of a L3 15 GeV E_T jet, as a function of the offline jet E_T (uncorrected).
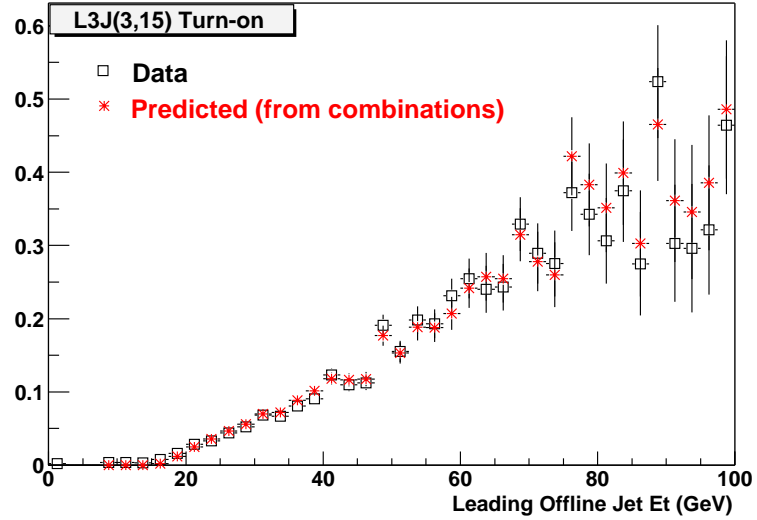
Figure 6.5: Comparison of the predicted probability of satisfying the L3J(3,15) L3 trigger and the actual probability, measured on the independent CJT (3,5) triggered data sample, as a function of the E_T (uncorrected) of the highest E_T (uncorrected) offline jet.

Table 6.1: Average trigger efficiencies for various signals and backgrounds. The two trigger versions are weighted by the integrated luminosity of data collected on each one. The trigger efficiencies are calculated after a basic set of offline analysis cuts: one jet with $E_T>$ 45 GeV, two jets with $E_T>$ 35 GeV, and at least four jets with $E_T>$ 15 GeV.

Process	Average trigger efficiency (%)
bh(100)	69
bbh(100)	68
bh(120)	73
bbh(120)	69
bh(150)	80
bbh(150)	76
$b\bar{b}$j	75
$b\bar{b}$jj	83
$b\bar{b}b\bar{b}$	68
Z($b\bar{b}$)+jets	64
$t\bar{t}$	93

6.1.3   Efficiency

The overall average efficiency of the two triggers used, weighted by the integrated luminosity of data collected with each, is shown in Table 6.1 for various signal and background Monte Carlo samples. The trigger efficiencies are calculated after a basic set of offline analysis cuts: one jet with $E_T>$ 45 GeV, two jets with $E_T>$ 35 GeV, and at least four jets with $E_T>$ 15 GeV. The trigger efficiencies after optimized analysis cuts are quoted in Table 8.2.

Rather than apply the same overall average trigger efficiency to each event in each sample, the probability for each event to pass the trigger is calculated, and the event is weighted by this probability. This method is more accurate since kinematical trigger biases are properly reproduced. Since two trigger versions were used to collect the
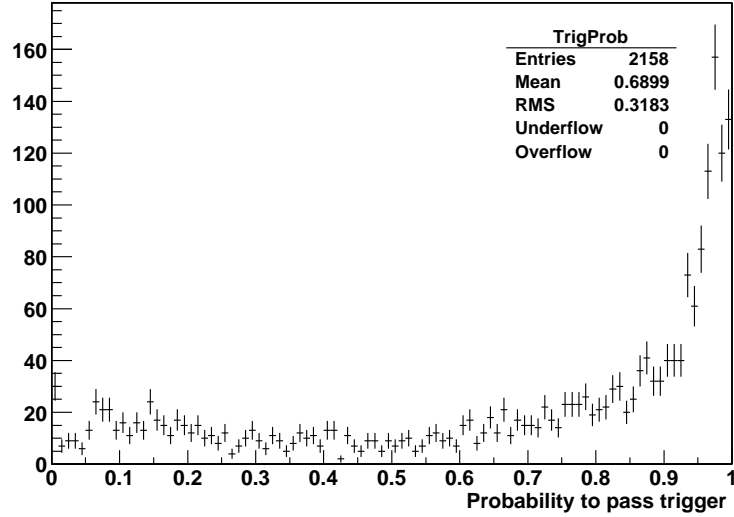
Figure 6.6: The probability calculated to pass the average trigger (weighted by integrated luminosity of data taken with each) for events in the bbh Monte Carlo sample with $m_h = 120$ GeV/c^2, after basic offline analysis cuts (see text).

data, with different integrated luminosities, both are taken into account with their respective pass probabilities weighted by their integrated luminosities. Figure 6.6 shows the distribution of the trigger pass probability for a typical simulated signal.

6.2 Data Sample

The data taken from September, 2002 – July, 2003 (131 pb^{-1}) is reconstructed with the p13.05.00 and p13.06.01 versions of the DØ reconstruction software. Of this sample, events are selected with one jet reconstructed with E_T>20 GeV and another two jets with E_T>15 GeV, in $|\eta|$<2.5 (un-corrected, and before any jet quality cuts). There are a total of 30.3 million events in this initial data sample.

Events are then processed through a custom software package, creating files containing just the quantities necessary for this analysis. Events from runs in which Calorimeter problems existed (such as intermittent noise, coherent noise, or too many

hot cells) are excluded. Approximately 15% of events in the selected data sample are from these bad runs. Only events which pass one of the two multi–jet triggers described are included in the analysis.

Since selection cuts on the jets' E_T are applied to the data before jet energy scale corrections, the same cuts must be simulated for the Monte Carlo events. The uncorrected jet E_T that would be measured in data must be accurately reproduced, for each simulated jet. One method is to first correct the MC jets with the MC jet energy scale factor, and then the inverse of the data jet energy scale factor, returning the simulated uncorrected jet E_T. Equivalently however, the selection cuts applied to the Monte Carlo jets can be adjusted, as opposed to recalculating the simulated uncorrected jet E_T for each Monte Carlo jet. The selection cut must be divided by the ratio of the MC JES factor to the data JES factor. For a jet E_T cut of 15 GeV, the MC JES factor is 1.3, and the data JES scale factor is 1.4. Their ratio is $1.3/1.4 = 0.93$. The 15 GeV jet E_T cut is then divided by this ratio, resulting in a cut of 16.1 GeV. The jet energy scale corrections are nearly identical for the 20 GeV cut on the leading E_T jet, resulting in a modified cut of 21.5 GeV. The systematic uncertainty introduced by this procedure, estimated by using the $\pm 1\sigma$ values of the jet energy scale corrections for both MC and data, is small ($\leq 2\%$). It is included in the acceptance uncertainty calculated for the jet energy scale in Chapter 7.

6.2.1 Data Quality Checks

The data is carefully examined for any excessively hot (noisy) calorimeter towers or trigger regions. Fixes for all known calorimeter problems significantly affecting the data quality were applied [11]. Figures 6.7 and 6.8 contain basic plots showing the quality and properties of the jets in the data sample. The n_j, E_T, and η spectra are sensible given the trigger, selection, and E_T cuts applied. There are no significant bumps at large jet E_T, η, or n_j, which would be signs of remaining detector problems. The $\eta - \phi$ distribution of the jets shows no sharp peaks, which would indicate remain-

ing calorimeter problems that have not been corrected for. The warm regions are most likely signs of remaining variation in the response of the calorimeter electronics to deposited energy (even after JES), hence more energy than average is measured there, and more jets in those regions will pass the jet E_T cuts.
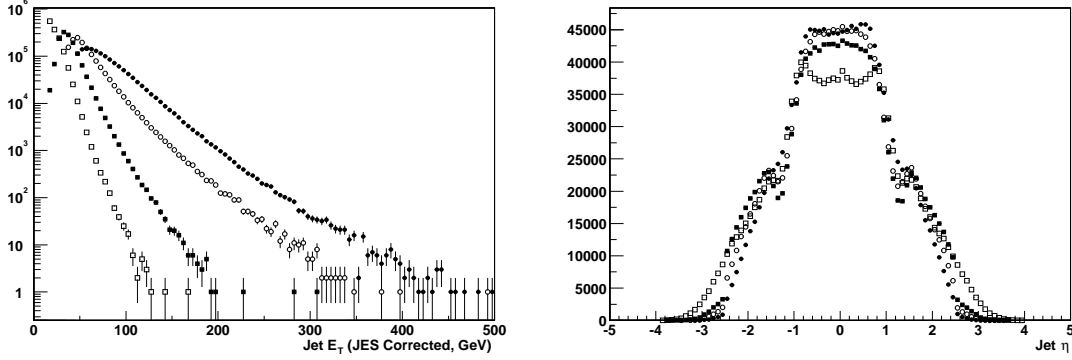
Figure 6.7: Left: The E_T distributions of the first (solid circles), second (open circles), third (solid squares), and fourth (open squares) highest E_T jets in each event. Right: The η distributions of the first (solid circles), second (open circles), third (solid squares), and fourth (open squares) highest E_T jets in each event. For both plots, jets must pass all quality cuts and taggability requirements, each event must pass one of the multi–jet triggers, jet E_T's are corrected for jet–energy scale, and the events must pass E_T cuts of 45, 35, and 15 GeV for the first, second, and third leading jets.
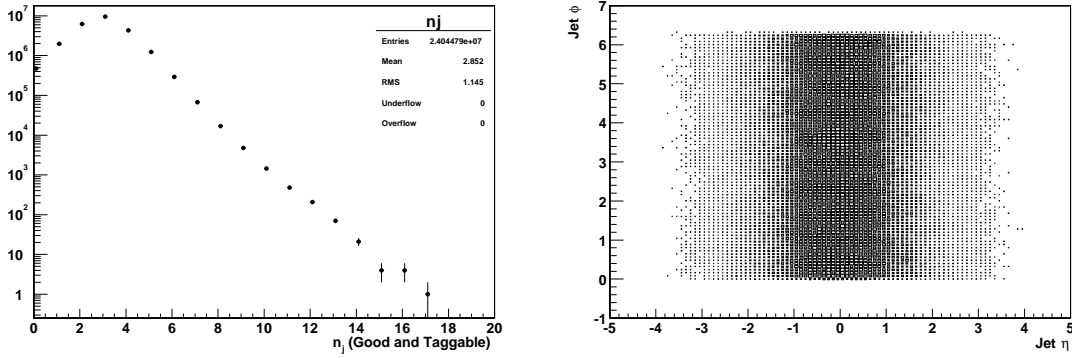
Figure 6.8: Left: The distribution of the number of jets in each event. Right: The $\eta - \phi$ of each jet in each event. For both plots, jets must pass all quality cuts and taggability requirements, each event must pass one of the multi–jet triggers, jet E_T's are corrected for jet–energy scale, and the events must pass E_T cuts of 45, 35, and 15 GeV for the first, second, and third leading jets.

To estimate the extent to which the jet selection criteria and taggability requirements remove hot–spots (noisy calorimeter regions) and fake jet contamination, the $\eta - \phi$ plot of jets is studied after combinations of criteria. Figure 6.10 shows the reconstructed jets before any quality cuts or taggability requirements, Figure 6.11 shows the jets which have passed quality cuts, Figure 6.12 shows the jets which have only passed taggability requirements, and Figure 6.9 shows the jets passing both the quality cuts and taggability requirements. Before any jet quality cuts or taggability requirements, there is a source of fake jets (from noisy calorimeter towers) around η of -0.6 and ϕ of 5.6 (see Figure 6.10). These jets are removed by the jet quality cuts alone, as seen in Figure 6.11. The fake jets in this region are also nearly removed by the taggability requirements alone, as seen in Figure 6.12. The region remains about 10% warmer than average, whereas it had been about 50% warmer than average before taggability requirements (judged from the color scale on the right of the figure). Hence the taggability seems to have a rejection factor of about 5 against fake jets. When combined with the jet quality cuts (which alone leave about 5% fake jets, see Chapter 4), less than 1% of the remaining jets are estimated to be fake. This fake rate has a negligible effect on the results, as shown in Chapter 7.

The remaining warm regions after both jet quality cuts and taggability requirements have at most 15% more jets than average regions. The higher number of jets in these regions can be caused by an energy scale shift of as little as 5%, due to the falling exponential E_T spectrum, as seen in Figure 6.7. This energy scale shift is smaller than the total uncertainty on the jet energy scale. Work is underway to measure the jet energy scale as a function of jet ϕ, which will reduce this error significantly in the future.
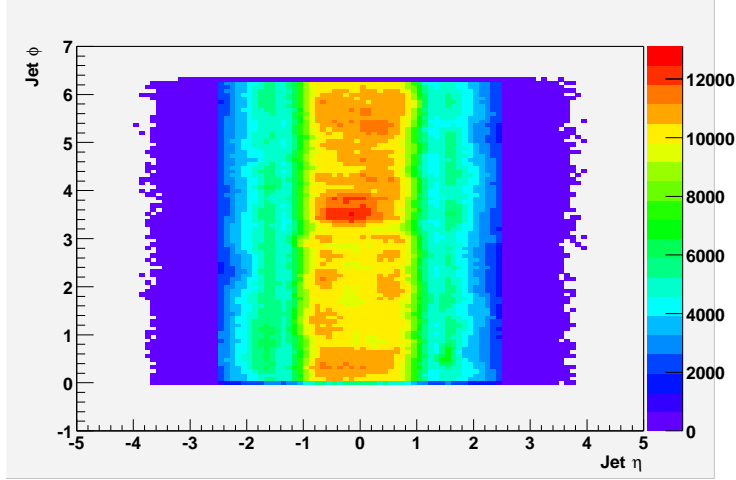
Figure 6.9: The $\eta - \phi$ of every jet which passes all quality cuts and taggability requirements. Each event must pass one of the multi–jet triggers, jet E_T's are corrected for jet–energy scale, and the events must pass E_T cuts of 45, 35, and 15 GeV for the first, second, and third leading jets.
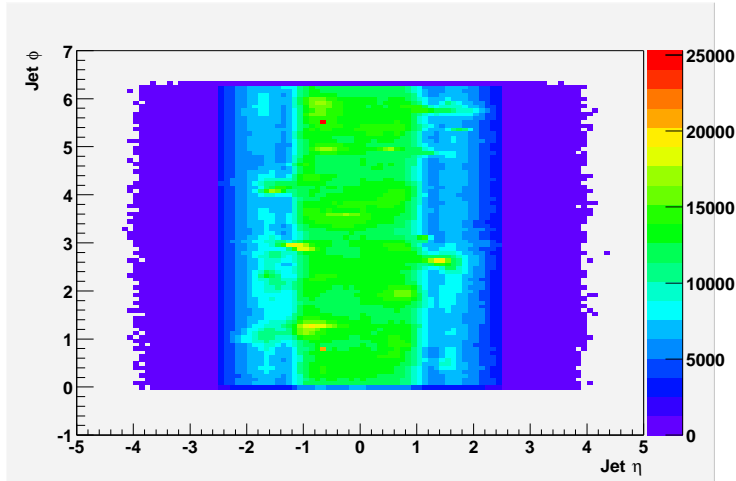
Figure 6.10: The $\eta - \phi$ of every jet in each event. Jets are not required to pass any quality cuts or taggability requirements. Each event must pass one of the multi–jet triggers, jet E_T's are corrected for jet–energy scale, and the events must pass E_T cuts of 45, 35, and 15 GeV for the first, second, and third leading jets.
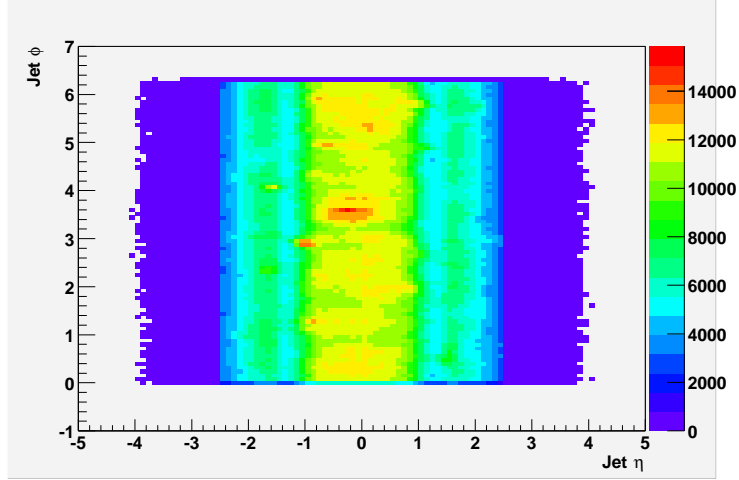
Figure 6.11: The $\eta - \phi$ of every jet which passes all quality cuts. Jets are not required to pass the taggability requirements. Each event must pass one of the multi–jet triggers, jet E_T's are corrected for jet–energy scale, and the events must pass E_T cuts of 45, 35, and 15 GeV for the first, second, and third leading jets.
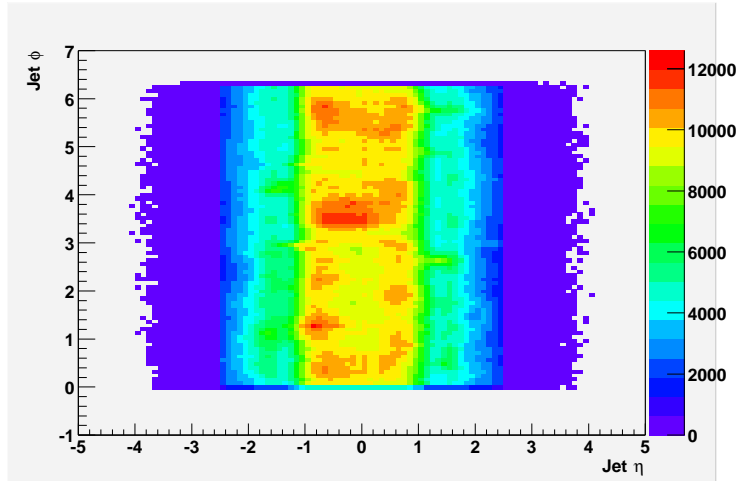
Figure 6.12: The $\eta - \phi$ of every jet which passes the taggability requirements. Jets are not required to pass any quality cuts. Each event must pass one of the multi–jet triggers, jet E_T's are corrected for jet–energy scale, and the events must pass E_T cuts of 45, 35, and 15 GeV for the first, second, and third leading jets.

Chapter 7

ANALYSIS

The signal from neutral Higgs bosons is expected to appear as an excess of events, with a particular invariant mass shape, in triple b–tagged multi–jet events. The Standard Model contributions to the triple b–tagged data must be understood to search for this signal. An optimal set of selection criteria are determined for isolating the expected Higgs boson signal from these backgrounds. The statistical methods employed for determining the signal exclusion limits are explained. Finally, the systematic uncertainties on the signal acceptance and background normalization are described, and their sizes are calculated.

7.1 *Background Determination*

Several Standard Model processes are responsible for the production of multi–jet events with three or more b–tagged jets. Multi–jet production from the Strong interaction is the largest source of background. The triple b–tagged multi–jet backgrounds are normalized to data, outside the signal region (around m_h), in the di–jet invariant mass distribution. Other significant backgrounds are the $Z(\rightarrow b\bar{b})$+jets and $t\bar{t}$ processes, whose contributions are calculated from Monte Carlo simulations.

Even though the multi–jet backgrounds are, in the end, calculated directly from data, they are first modeled in Monte Carlo and compared in detail to data. This provides an important cross-check, for both the interpretations of the data and the simulations of the SM. The comparison's agreement adds confidence to the understanding of the detector performance and data reconstruction methods. In addition,

the relative contributions to the production of multi–jet events containing heavy–flavor (bottom and charm quarks) are understood. This is interesting from a theoretical perspective, and is also useful for designing future searches at the Tevatron and other experiments planned for the future.

7.1.1   Monte Carlo Cross-Checks

Calculations of multi–jet production processes are not available at NLO. Even the full LO simulations, which have just recently become available (as discussed in Chapter 5), suffer large theoretical uncertainties (from the choice of factorization scale, for instance). Also, the fraction of heavy–flavor (containing c– or b–jets in the final state) in multi–jet events is difficult to simulate accurately. For these reasons, the simulations of multi–jet events must themselves be tested in as many ways as possible.

The double b–tagged data (events with two or more b–tagged jets) are first compared to the Monte Carlo. This sample provides a high–statistics arena for checking the accuracy of the simulations and the methods used for making the comparison. The invariant mass distribution of the leading two E_T jets in the double b–tagged data is fit to a sum of backgrounds, as seen in Figure 7.5. Good agreement with the data is seen, both for the shape of the distribution and the normalization of the heavy–flavor processes predicted by the simulation.

Next, the triple b–tagged data is compared to Monte Carlo, using what was learned from the comparison of the double b–tagged data. The invariant mass distribution of the leading two E_T jets is again fit to a sum of backgrounds. Good agreement between the simulation and the data remains.

Monte Carlo Cross-Checks of Double b–Tagged Data

The double b–tagged multi–jet background is divided into three categories, for the purpose of accounting and for comparison with the simulations:

- *Multi–jet fakes* is light multi–jet production. Two or more jets are light–quark or gluon jets that have been falsely identified as b–jets. Events with a final state gluon that has split into a pair of nearly collinear b– or c–quarks ("gluon–splitting") are also included in this category, if the split gluon jet is b–tagged.

- *Heavy–flavor (HF)* multi–jet production, $b\bar{b}$j(j), has two real b–jets, both of which have been b–tagged. The contribution of the $c\bar{c}$j(j) process is studied below.

- *Other backgrounds*, such as $b\bar{b}b\bar{b}$, Z→$b\bar{b}$+jets, and $t\bar{t}$ make up the final category.

The expected signal contribution to the double b–tagged data is negligible and does not affect the normalization of the backgrounds.

The multi-jet fakes are estimated from data. Using the full data sample (events with zero or more b–tags), the probability of b–tagging a jet is measured, as a function of the E_T of the jet, in three different $|\eta|$ bins [1], as shown in Figure 7.1. This set of functions is called the *fake–tag parameterization*, although it is understood to have some contamination at this point from true HF events (containing pairs of well–separated c– or b–jets) in the data sample from which it was derived. This HF contamination will be corrected for, by a procedure described below [2]. The fake–tag parameterization is used to estimate the multi–jet fakes contribution by applying it to every jet in the full sample. The probability for an event to have two or more fake b–tags is the probability for that event to enter the multi–jet fakes distribution.

The b–tagging used in this analysis is unable to distinguish contributions from bottom and charm quark events. However, the efficiency for tagging a charm–jet is about 1/4 of that for tagging a bottom–jet, as determined from MC simulations,

---

[1]The fraction of jets tagged changes slightly as a function of $|\eta|$. It is higher in the central region, and lower in the forward regions, as seen in Figure 7.1.

[2]The contamination from gluon–jets where the gluon has split into a nearly collinear $b\bar{b}$ or $c\bar{c}$ pair can not be corrected for.

as shown in Figure 4.24. Therefore, when two b–tags are required, as is the case in the fit of the double b–tagged data, the fraction of $c\bar{c}$j(j) events relative to $b\bar{b}$j(j) events will be a factor of $\sim 4^2 = 16$ times lower after b–tagging than it was before. The fraction of $c\bar{c}$jj to $b\bar{b}$jj prior to b–tagging is estimated using the MADGRAPH Monte Carlo generator. Using the same generator–level cuts as for the ALPGEN and MADGRAPH $b\bar{b}$jj Monte Carlo generated (listed in Table 5.2), the $c\bar{c}$jj cross–section was estimated to be 3240±174 pb, only 22% higher than the $b\bar{b}$jj cross–section estimated with MADGRAPH. The contribution of $c\bar{c}$j(j) in the double b–tagged data sample is estimated to be $\sim 1.22/16 = 8\%$ of the events. Thus, the $b\bar{b}$jj normalization is understood to contain a small contribution, approximately 8%, from the $c\bar{c}$jj process.

The normalization of the HF multi–jet contribution is left as a free parameter in the fit of the double b–tagged data. The contribution of the $b\bar{b}b\bar{b}$ background is expected to be small compared to the HF multi–jet background in the double–tagged sample, thus its normalization will have little effect on the result of the fit. The $b\bar{b}b\bar{b}$ normalization is fixed to the ratio of HF multi–jet production to $b\bar{b}b\bar{b}$ production as given by ALPGEN. The normalization of the Z($\rightarrow b\bar{b}$)+jets and $t\bar{t}$ backgrounds are fixed to their accepted values.

The result of the fit of the double b–tagged data to backgrounds is shown in Figure 7.2. The shape of the invariant mass distribution is well modeled. The cross–section for the HF multi–jet processes predicted by ALPGEN must be multiplied by a factor of only 0.98 ± 0.02 to agree with data, in very good agreement.

Now that an estimate has been made of the HF content of the double b–tagged data sample, the fraction of the fake–tag parameterization that was due to these events is estimated and subtracted from the fake–tag parameterization. The fraction of HF events in the full data sample is simply the ratio of the number of non–tagged HF multi–jet events, as estimated from the ALPGEN sample normalized using the above fit, to the number of non–tagged events observed for the total data sample. Given that there are about 15,600 ALPGEN HF events before b–tagging (using the

116
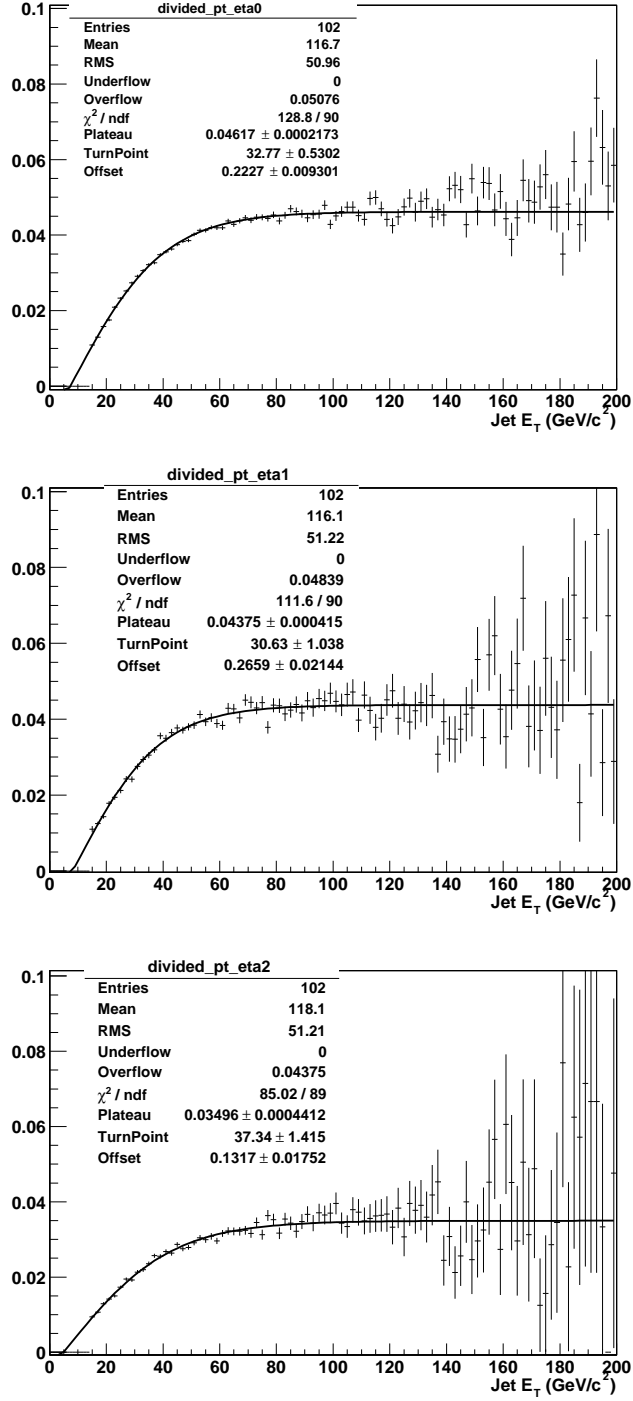
Figure 7.1: The fake–tag parameterization (a set of *tanh* functions), which is the probability of tagging a jet in the inclusive data sample. The parameterization, a function of jet E_T, is divided into three bins: $|\eta|<1.1$ (top), $1.1 <|\eta|<1.5$ (center), and $1.5 <|\eta|$ (bottom).
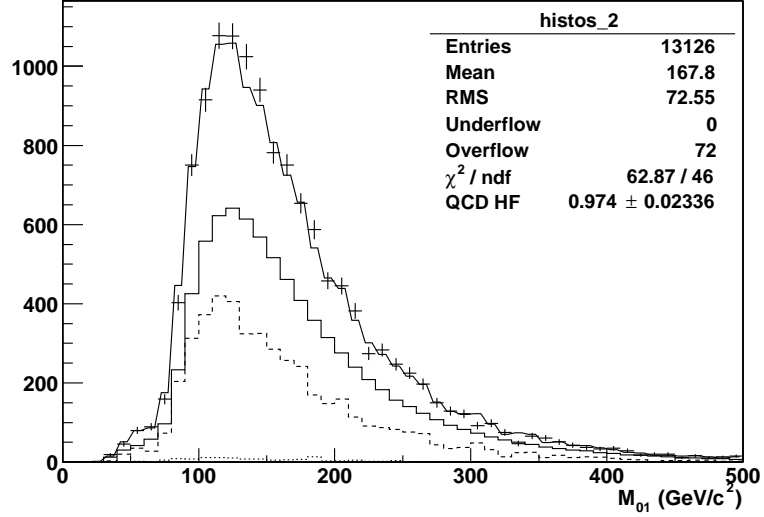
Figure 7.2: Initial fit of the double b–tagged data's two leading jet invariant mass spectrum to a sum of backgrounds: multi–jet fakes (solid), ALPGEN $b\bar{b}$jj MC (dashed), and other small backgrounds (Z($\rightarrow b\bar{b}$)+jets, $t\bar{t}$, and ALPGEN $b\bar{b}b\bar{b}$) (dotted).

normalization derived above) and 1.34 million events in the full data sample, the HF fraction is estimated to be 1.2% using this method. The probability of tagging a jet in the ALPGEN HF multi–jet sample, as a function of the E_T of the jet, is shown in Figure 7.4. The fake–tag parameterization can be corrected by subtracting the tag–rate for the ALPGEN HF sample (29%) times the HF fraction (1.2%), equal to a reduction in tag–rate probability of 0.35%. This correction lowers the plateau of the average fake–tag parameterization (see Figure 7.3) from 4.4% down to 4.05%.

However, the fraction of observed HF after the fake–tag parameterization correction would then be larger since there would be slightly less multi–jet fakes, and then the fake–tag parameterization would be re–done with this new HF fraction, ad infinitum. Therefore, it is assumed that the first correction measured, $c = (29\%/4.4\%) \times 1.2\%$ $= 7.5\%$, is the first term in a geometric series, so the total correction factor is $c/(1-c)$ $= 8.1\%$. The fit of the double b–tagged data is now re–done, with this final fake–tag

118

histos_50	
Entries	102
Mean	120.6
RMS	52.25
Underflow	0
Overflow	0.04274
χ^2 / ndf	110.4 / 66
Plateau	0.04431 ± 0.0001903
TurnPoint	34.86 ± 0.4947
Offset	0.2 ± 0.007687
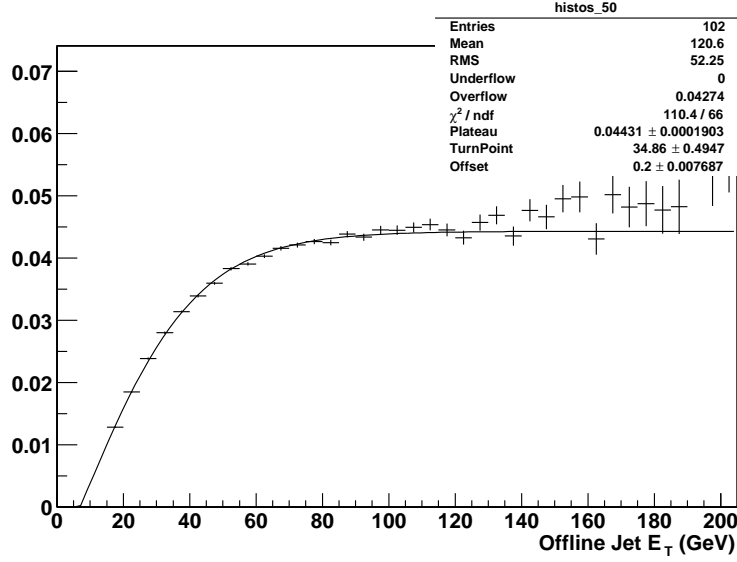
Figure 7.3: The b–tag parameterization (a tanh function) of the probability to b–tag a jet in the full data sample.

parameterization (lowered by the correction factor, 8.1%), and is shown in Figure 7.5. After corrections, the HF multi–jet processes are a factor of 1.22 ± 0.02 higher in data than predicted by ALPGEN. This disagreement is smaller than the theoretical uncertainties associated with the multi–jet cross–section calculation.

Monte Carlo Cross-Checks of Triple b–Tagged Data

Using the normalization for multi–jet background processes derived above from the double b–tagged sample, the background rates and invariant mass distribution for the triple b–tagged data sample can now be checked in Monte Carlo. The background events to the triple b–tagged data are categorized, in much the same way as the double b–tagged backgrounds:

- Multi–jet fakes, where three (or more) light jets have been falsely b–tagged.

- HF multi–jet production, $b\bar{b}$j(j), where two of the jets are real b–jets but one
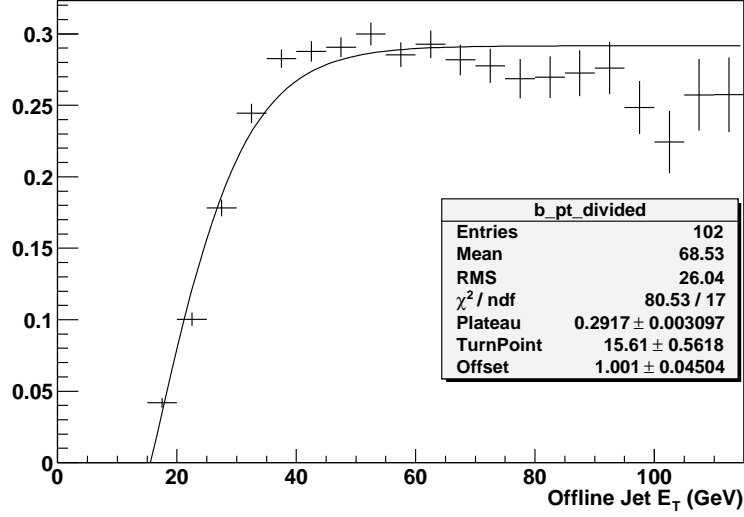
Figure 7.4: The b–tag parameterization (a tanh function) of the probability to b–tag a jet in the ALPGEN $b\bar{b}$jj Monte Carlo sample. (The MC b–tagging has been adjusted, using the scale factor to data, to reproduce the b–tagging expected for $b\bar{b}$jj events in data.)

(or more) of the light–jets has been falsely b–tagged.

- Other backgrounds, such as $b\bar{b}b\bar{b}$, Z($\rightarrow b\bar{b}$)+jets, and $t\bar{t}$ production. Three or more of some combination of light– and b–jets are b–tagged.

The multi–jet fakes are once again estimated by applying the corrected fake–tag parameterization, but now using the probability that three or more jets are b–tagged by the function. The HF multi–jet contribution is modeled from the ALPGEN HF multi–jet Monte Carlo ($b\bar{b}$j(j)), using the normalization derived above from the double b–tagged sample. The b–tagging is applied in the simulation, and those events with three or more b–tags (one or more of which will be a false b–tag) are kept. Similarly, the $b\bar{b}b\bar{b}$ contribution is taken from the ALPGEN Monte Carlo, using the same normalization factor. The Z($\rightarrow b\bar{b}$)+jets and $t\bar{t}$ backgrounds are also taken from simulation, keeping those events which have three or more b–tags (some of which

120
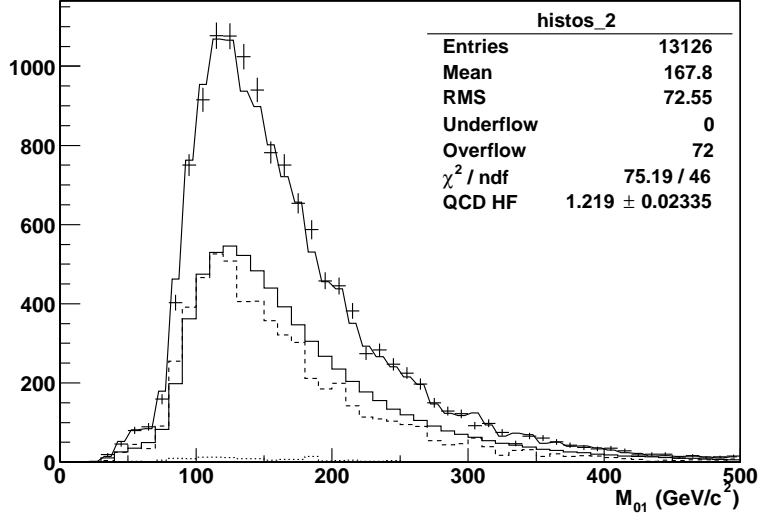
Figure 7.5: Final fit of the double b–tagged data's two leading jet invariant mass spectrum, after fake–tag parameterization correction, to a sum of backgrounds: multi–jet fakes from data (solid), ALPGEN $b\bar{b}$jj MC (dashed), and other small backgrounds ($Z(\rightarrow b\bar{b})$+jets, $t\bar{t}$, and ALPGEN $b\bar{b}b\bar{b}$) (dotted).

must be falsely identified light–jets), but their normalization are still fixed to their accepted values. The invariant mass spectrum of the leading two jets in the triple b–tagged data sample is shown in Figure 7.6, with the contributions from the various backgrounds shown.

7.1.2 Triple b–Tagged Background Calculation

The multi–jet background to the triple b–tagged data is now calculated directly from data. The corrected fake–tag parameterization is applied to the non–b–tagged jets in each event in the double b–tagged data sample, and the probability is calculated for each event to have three or more b–tags (including the contribution from both the real tags and those estimated from the fake–tag parameterization). This estimation of the background will be referred to as the *triple b–tag estimation*. This of course neglects the contribution from processes which have more than two real b–jets, such
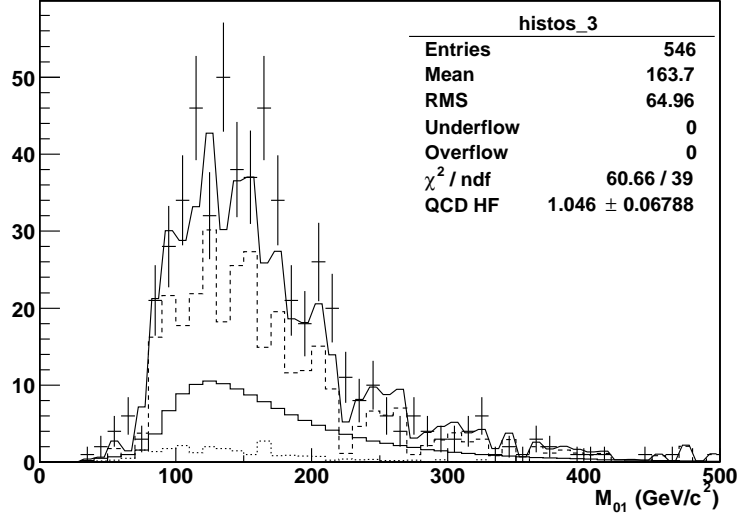
Figure 7.6: Fit of the triple b–tagged data's two leading jet invariant mass spectrum, after fake–tag parameterization correction, to a sum of backgrounds: multi–jet fakes (solid), ALPGEN $b\bar{b}$jj MC (dashed), and other small backgrounds (Z($\rightarrow b\bar{b}$)+jets, $t\bar{t}$, and ALPGEN $b\bar{b}b\bar{b}$) (dotted).

as $b\bar{b}b\bar{b}$ and Z($\rightarrow b\bar{b}$)+$b\bar{b}$. However, the shape of these backgrounds is very similar to the double b–tagged spectrum, since it shares a similar production mechanism. The invariant mass spectrum of $b\bar{b}b\bar{b}$ is similar to that of $b\bar{b}$jj and that of Z($\rightarrow b\bar{b}$)+$b\bar{b}$ is similar to Z($\rightarrow b\bar{b}$)+jets, as determined from Monte Carlo simulations.

The overall background normalization can be determined by fitting the triple b–tagged invariant mass distribution outside the signal region ($\pm 1\sigma$ of the Gaussian fit to the signal) with the triple b–tag estimation. The resulting fit using this method of background normalization is shown in Figure 7.7 for $m_h = 120$ GeV/c^2. This is the final method used for setting limits, in Chapter 8.

Alternatively, the triple b–tagged data can be fit to a sum of the signal Monte Carlo shape and the triple b–tag estimation. The resulting fit using this latter method is shown in Figure 7.8 for a signal mass of $m_h = 120$ GeV/c^2.
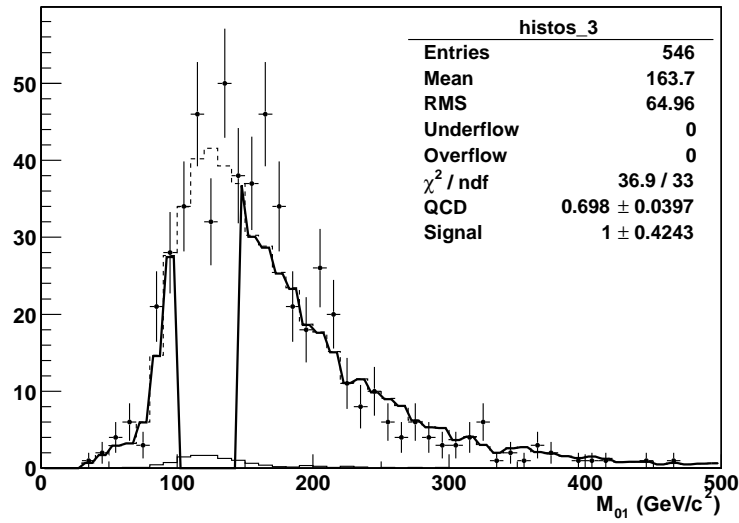
Figure 7.7: Fit of the triple b–tagged data's two leading jet invariant mass spectrum (points) to the triple b–tag estimation function outside the signal region (shown as vertical lines), for $m_h = 120$ GeV/c^2. Also shown are the signal expected for $\tan\beta$ of 50 (thin solid) and a fit to the triple b–tag estimation function without excluding the signal region (dashed histogram).
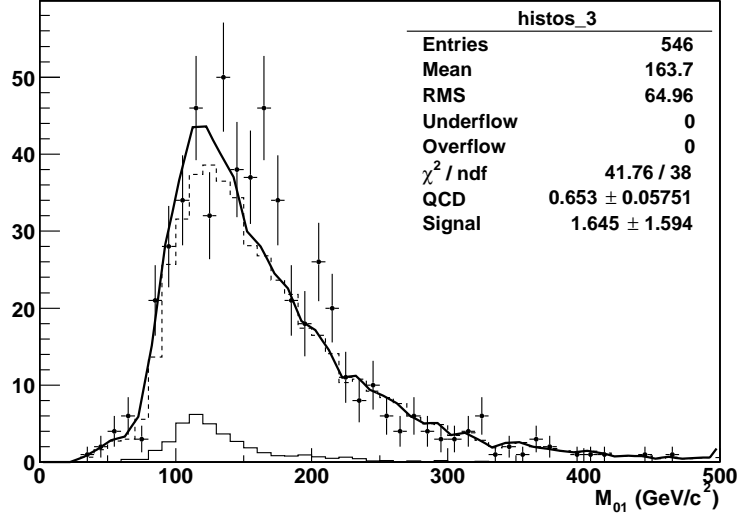
Figure 7.8: Fit of the triple b–tagged data's two leading jet invariant mass spectrum (points) to a sum of the triple b–tag estimation function (dashed) and signal for m_h = 120 GeV/c^2. The sum is shown in thick solid and the signal, scaled to the fit, is in thin solid.

7.2   Analysis Methods

The analysis cuts are optimized by comparing the distributions of simulated signal events remaining after the cuts to the simulated backgrounds. Cut values are chosen which give the best predicted 95% C.L. exclusion limit on associated Higgs production, for each m_h.

Limits on $\tan\beta$ are calculated using two methods, and their results are compared. The first method fits the signal's invariant mass distribution to a Gaussian distribution and uses this shape to weight the signal and background events. The second method uses the full shapes of the signal and background distributions to weight events without fitting to a parameterized model. This second method is almost always found to be more powerful, since it uses more information about the shapes. The first method is useful as a cross-check, due to its relative simplicity.

7.2.1 Limit Setting

For a given set of analysis cuts, the invariant mass of the leading two E_T jets in the triple b–tagged data sample is examined. The background is calculated via the method described above: fitting the triple b–tag estimation's invariant mass distribution to the triple b–tagged data outside the signal region. The signal distribution (for $\tan\beta = 50$) is fit to a Gaussian distribution, and $\pm 1\sigma$ is used to define the signal region. An example for m_h=120 GeV/c^2 is shown in Figure 7.7.

The Gaussian Weighting Method

This method uses the Gaussian distribution fit to the signal to weight each di-jet invariant mass bin. Bins in which more signal is expected are given larger weights. The weight given to each bin is proportional to the height of the Gaussian distribution at the center of the bin. The weights are normalized such that the largest is unity, and the smallest weight is zero. This method partially takes advantage of the expected shape of the signal distribution. It is more sensitive than a simple square–cut method, which gives zero weight to bins outside the signal region and weights of unity to those bins inside the signal region.

Table 8.1 shows the means and widths of the fits to the signal distributions for all m_h studied, using the optimized set of cuts determined below. The total number of weighted data and predicted background events is counted, for each trial m_h signal distribution. The total numbers of weighted events are shown in Table 8.3.

The results of each *counting experiment* are translated into 95% Confidence Level (C.L.) limits on $\tan\beta$, using the standard Bayesian technique recommended for DØ [38]. The model assumes that the average number of observed events, μ, will be:

$$\mu = b + \pounds\epsilon\sigma \tag{7.1}$$

where b is the expected number of background events, $\pounds$ is the integrated luminosity,

ϵ is the signal efficiency, and σ is the signal production cross-section. The probability of observing k events, given a mean μ is assumed to follow Poisson statistics:

$$P(k|\mu) = \frac{e^{-\mu}\mu^k}{k!} \tag{7.2}$$

A flat prior probability density is assumed for the cross-section, from 0 to σ_{max}. Tests are performed to test that resulting limits are independent of the σ_{max} assumed. Bayes' theorem is applied, and the value of the upper limit on signal production cross-section, σ_{UL}, is found which maximizes the inequality:

$$C.L. > \int_0^{\sigma_{UL}} d\sigma \rho(\sigma|k) \tag{7.3}$$

where a 95% C.L. is required. ρ is the probability density derived from $P(k|\mu)$ using Bayes' theorem and includes the effects of systematic uncertainties on b, ϵ, and $\pounds$.

The Limit 1 column of Table 8.3 shows the 95% C.L. limits derived using the Gaussian weighting method.

The Full Signal Shape Method

The Monte Carlo signal shapes of the invariant mass of the two leading E_T jets in the triple b–tagged events are not perfectly Gaussian. Also, the expected background shape is not flat. For these reasons, maximal sensitivity to signal production is not obtained by using the Gaussian weighting limit setting method.

The MCLimit method [39] was designed to interpret the data for MSSM Higgs searches at LEP2. It derives limits on signal production, directly using the shapes of the hypothesized signal and expected background invariant mass distributions. Each histogram bin is treated as an independent experiment, and their significance are combined. For each invariant mass bin, the *likelihood ratio*, X_i, is given by Poisson

statistics to be:

$$X_i = \frac{e^{-(s_i+b_i)}(s_i+b_i)^{d_i}}{d_i!} \bigg/ \frac{e^{-b_i}b_i^{d_i}}{d_i!} \tag{7.4}$$

where s_i, b_i, and d_i are the number of signal, expected background, and observed events, respectively. The *combined likelihood ratio* for a data set, X_d, is the product of all the likelihood ratios of the bins:

$$X_d = \prod_{i=1}^{n} X_i \tag{7.5}$$

The Confidence Level for signal plus background production (CL_{s+b}) is computed as a sum over all possible experiments, e, which have $X_e < X_d$, where X_e is the combined likelihood ratio for the experiment. The variable summed over is the combined likelihood (not the ratio) for signal plus background:

$$CL_{s+b} = \sum_{X_e < X_d} \prod_{i=1}^{n} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{e_i}}{e_i!} \tag{7.6}$$

where e_i is the number of events observed in bin i of experiment e. The sum is computed using Probability Distribution Function (PDF) techniques and numerical Monte Carlo methods.

The C.L. for the background-only hypothesis (CL_b) is calculated in the same way as CL_{s+b}. The C.L. for signal (CL_s) is then derived using the Modified Frequentist approach:

$$CL_s = CL_{s+b}/CL_b \tag{7.7}$$

The CL_s is used to derive the final limits on signal production.

The effects of systematic uncertainties are taken into account when computing the sums over the combined likelihoods. The possible values of the signal and background in each bin are averaged over, using the input systematic uncertainty distributions [3].

---

[3] The systematic errors are assumed to smear the number of events according to Gaussian distributions.

Table 7.1: Loose initial analysis cuts (to be tightened by optimization).

Variable	Cut value
E_T's of all jets (JES corrected)	>15 GeV
$|\eta|$'s of all jets	<3
n_j	≤ 6

The systematic errors are also averaged over when calculating the combined likelihood ratios, X_d and each X_e.

The implementation of the MCLimit routines in ROOT is used [4]. The signal histogram, derived assuming $\tan\beta=1$, is scaled by $\tan^2\beta$ at each trial $\tan\beta$ value. The value of $\tan\beta$ is varied, starting at 50, either up or down until the CL_s obtained is <5%, but as close to 5% as possible. The results using this limit setting method are shown in the Limit 2 column of Table 8.3.

7.2.2 Optimization

The selection cuts used in the analysis are optimized using Monte Carlo. It was verified above that the shapes of the simulated backgrounds accurately model both the double and triple b–tagged data. Thus it is safe to assume that an approximately optimal set of analysis cuts can be achieved by using the simulated backgrounds alone. There is no choice but to use simulated signal events.

The quantity to be optimized is the expected signal production limit, as calculated using the methods discussed above. Systematic errors are not included in the optimization procedure. A separate optimization is performed for two values of n_j^{min}, the minimum number of jets allowed in the final state, 3 and 4. The case when $n_j^{min}=3$ may be referred to as the "bh" search and $n_j^{min}=4$ as the "bbh" search. Beginning with a basic and loose set of analysis cuts, shown in Table 7.1, one parameter is opti-

---

[4]ROOT version 3.05/07 is used in this analysis.

Table 7.2: Optimized analysis cuts for each m_h and for both the n_j^{min} cases (3 and 4).

Signal (GeV/c^2)	E_T1(GeV)	E_T2(GeV)	E_T3(GeV)	$|\eta|_j$	n_j^{max}
n_j^{min}=4, m_h=100	40	35	15	2.5	4
n_j^{min}=4, m_h=120	45	35	15	2.5	5
n_j^{min}=4, m_h=150	60	40	15	2.0	5
n_j^{min}=3, m_h=100	45	35	15	2.5	4
n_j^{min}=3, m_h=120	45	35	15	2.5	4
n_j^{min}=3, m_h=150	60	40	15	2.0	4

mized at a time, by varying the parameter, deriving the expected 95% C.L. limit, and choosing the value of the parameter which minimizes the expected signal production limit. The parameters are also optimized separately for each m_h studied (in addition to the two n_j^{min}). The parameters studied were (in order of their optimization):

- The minimum E_T of the leading, second–leading, and third–leading jet

- The maximum $|\eta|$ of the jets

- n_j^{max}, the maximum number of jets allowed ($3-5$)

The resulting optimized values for each m_h for both the n_j^{min} cases are shown in Table 7.2. The optimal cuts are roughly independent of the Higgs mass and the number of jets allowed. Harder E_T cuts of 60 and 40 GeV are used for the leading and next–to–leading jets and slightly tighter η cuts, for m_h=150 GeV/c^2. Also, 5 jet final states are preferred in the n_j^{min}=4 channel for all but the lightest Higgs mass studied. No events were considered with more than 5 jets in the n_j^{min}=4 channel, or 4 jets in the n_j^{min}=3 channel, since the signal Monte Carlo, which relies on Pythia radiation to simulate "extra" jets, is not considered reliable beyond one extra jet.

7.3 Systematic Errors

The sources of systematic uncertainty fall into two general categories: errors affecting the calculation of signal acceptance, and errors that affect the normalization of the number of estimated background events. The final errors assigned for the signal acceptance are shown in Table 7.4, and the background normalization errors used are shown in Table 7.5.

7.3.1 Acceptance Systematics

The errors from all sources which affect the signal acceptance will be added in quadrature, and are shown in Table 7.4.

The absolute value of the integrated luminosity that the collected data sample corresponds to has an uncertainty of 10% [40]. The error is dominated by the uncertainty in the far forward inelastic $p\bar{p}$ scattering cross–section, and by luminosity monitor calibration uncertainties. The luminosity uncertainty is accounted for by the limit calculation process, which deduces the cross–section limit from the observed number of data and expected background events.

A major source of acceptance uncertainty arises from a lack of perfect understanding of the signal generation process due to the limited order in perturbation theory at which Monte Carlo signal events are generated (LO) and at which total cross–sections are calculated (NLO). Simulated LO events are normalized to the NLO cross–sections and p_T spectra. Since this procedure is only an approximation to reproducing the full NLO kinematics of the events, an error of half the correction is assigned.

The remaining source of theoretical error is the extent to which the NLO kinematics and total cross–sections could differ from physical reality, which of course is not limited to a low order of perturbation theory. The magnitude of this error can be estimated by varying various parameters of the NLO calculation introduced by the perturbative methods, such as the factorization and renormalization scales, the run-

ning value of the bottom–quark mass, and the renormalization scheme (such as OS or $\overline{MS}$). Based on information from the NLO calculations of the bh [18] and bbh signals [16] [17], theoretical uncertainties were estimated. Reference [18] shows that the total uncertainty on the bh cross–section is about 5%. The cross–section of bbh shows much more variation at NLO, up to 20%, when the factorization/renormalization scale is moved from $m_h/4$ to m_h, as seen in Reference [16], or when the renormalization scheme is changed as shown in Reference [17]. These further theoretical errors are not included in the analysis; the quoted central values are used instead.

The overall trigger efficiency is subject to errors coming from the limited statistics of the data samples used to measure the trigger efficiencies, the inaccuracy with which the parameterized turn–on curves represent the true turn–ons, and the limitations of the assumptions made about the independence of jets at the trigger level. The error from limited statistics can be estimated from the errors on the fit parameters of the turn–on curves to the data. Typical fit parameter errors are at the level of 2% and never exceed 5%. A conservative error of 4% is chosen for each turn–on curve. The error due to the inability of the fitting function to perfectly model the data can be estimated from the χ^2/NDF values of the resulting fits. Typical χ^2/NDF values are <1.1, and are at their largest <2. The total error will be multiplied by a conservative $\sqrt{\chi^2/NDF}$ value of 1.5. Finally, the inaccuracies involved in the isolated jet hypothesis made to combine turn–ons from individual jets into event probabilities is estimated by how well the closure–tests reproduce the observed trigger rates on the independent data samples. Typical errors involved in the closure–tests (the average difference between the predicted trigger rate and the observed trigger rate) are about 5%. The total trigger uncertainty is thus estimated to be ±9%.

The Gaussian widths of the signal peaks for each m_h are derived from simulation. If the jet energy resolution is different in data than in the Monte Carlo, more or less signal events would be included under the Gaussian peak in reality than is assumed. The measured uncertainties for the Monte Carlo and data jet energy resolutions,

Table 7.3: The error on the width of the Gaussian fit to each signal and the final signal acceptance error, for each m_h and minimum number of jets allowed.

Signal (GeV/c^2)	Resolution Error (GeV)	Acceptance Error (%)
$n_j^{min}=4$, $m_h=100$	2.2	8.0
$n_j^{min}=4$, $m_h=120$	2.8	12.0
$n_j^{min}=4$, $m_h=150$	3.2	11.9
$n_j^{min}=3$, $m_h=100$	2.1	7.5
$n_j^{min}=3$, $m_h=120$	2.4	12.5
$n_j^{min}=3$, $m_h=150$	3.0	12.8

shown in Figure 4.17, are added in quadrature. The resulting uncertainty on the difference between data and Monte Carlo resolution is 8% for jets with E_T between 40 and 60 GeV, averaged over the η ranges, weighted by the largely central signal jet η spectrum. Since the width of the signal distributions is approximately $\sqrt{2} \cdot \sigma$, where σ is the jet energy resolution, the width of the fitted signal distributions will be assigned an uncertainty of $\sqrt{2}$ times the jet energy resolution uncertainty, or 12%. These uncertainties in the widths of the signal distributions can be directly translated into uncertainties in signal acceptances, by recounting the number of weighted signal events using the narrower and wider Gaussian distributions, and taking the difference. Table 7.3 shows the errors in signal acceptance caused by uncertainty in jet energy resolution for each m_h.

The uncertainty in the jet energy scale (both statistical and systematic) is reported by the JES group, for each corrected jet. To calculate the uncertainty in the result as an effect of this scale uncertainty, the entire analysis has been performed by taking the upper value of the jet energy scale and again with the lower value (i.e. with the $\pm 1\sigma$ jet energy scale factors). The overall acceptances changed by the amounts shown in Table 7.4 in the Energy Scale column, which shows the dependence on m_h.

The extent to which the jet energy resolution error was correlated with the jet

energy scale error was studied. The change in the signal peak resolution was observed as the jet energy scale was varied by $\pm 1\sigma$. The width of the Gaussian signal peak decreased in proportion to the central signal peak value. The change in signal acceptance for the wider and narrower signal peaks was within 1% of the central value shown in Table 7.3, as the JES was varied by $\pm 1\sigma$. Thus the correlation between the error on the signal peak widths and the JES error is negligible, and the two errors are simply added in quadrature.

The jet reconstruction efficiency has been measured both in Monte Carlo and in data [41], as a function of E_T. Each jet in the simulation was adjusted for the difference between the two efficiencies, but the error on that difference correction remains. The error is estimated by taking the $\pm 1\sigma$ values for the jet reconstruction efficiencies (for both Monte Carlo and data) and running the full analysis. The accepted signal varied by the amounts shown in Table 7.4 in the Jet ID column, which shows the dependence on m_h.

The uncertainty on the b–tagging efficiency in data is represented in the error on the data to Monte Carlo scale factor, which is either a fixed value or parameterized as a function of jet E_T and $|\eta|$, depending on the algorithm. The analysis was run using the $\pm 1\sigma$ scale factors, and the acceptance varied by the amounts shown in Table 7.4 in the B–Tag column, which shows the dependence on m_h.

7.3.2 Background Systematics

There is a statistical error associated with the uncertainty in the normalization of the background, as fit outside the signal region. There is additional systematic uncertainty from the shape of the distribution not being modeled perfectly, which can be estimated from the χ^2/NDF of the background normalization fit. The statistical error would be multiplied by this measured $\sqrt{\chi^2/NDF}$ for each m_h.

Another method of calculating the uncertainty in the background normalization is to use the statistical uncertainty of the fake–tag parameterization fit multiplied by its

Table 7.4: The errors from each source (in percent), which are added in quadrature to give the total errors on acceptance.

Signal (GeV/c^2)	NLO/LO	Trig	Resolution	JES	Jet ID	B–tag	Total
n_j^{min}=4, m_h=100	5	9	8.0	20	3.8	13.5	27.7
n_j^{min}=4, m_h=120	5	9	12.0	16	3.4	13.5	26.5
n_j^{min}=4, m_h=150	5	9	11.9	13	3.5	13.8	24.9
n_j^{min}=3, m_h=100	5	9	7.5	12	3.7	13.5	22.4
n_j^{min}=3, m_h=120	5	9	12.5	7.5	3.5	13.2	22.5
n_j^{min}=3, m_h=150	5	9	12.8	3.4	3.6	13.4	21.8

Table 7.5: The errors on the background normalization, measured via both of the methods investigated, and the total error assigned, for each m_h and number of minimum jets allowed.

Signal (GeV/c^2)	3 b–tag Fit (%)	Fake–tag Fit (%)	Total Error (%)
n_j^{min}=4, m_h=100	8.2	8	11.5
n_j^{min}=4, m_h=120	7.1	8	10.7
n_j^{min}=4, m_h=150	6.7	8	10.0
n_j^{min}=3, m_h=100	7.8	7	10.0
n_j^{min}=3, m_h=120	6.9	7	9.8
n_j^{min}=3, m_h=150	6.6	7	9.6

$\sqrt{\chi^2/NDF}$, since this function is used to propagate the shape of the double b–tagged data to the triple b–tagged data.

The results of both methods and the total errors used for the background normalization are listed in Table 7.5, as a function of m_h.

Chapter 8

RESULTS AND DISCUSSION

Upper limits on $\tan\beta$ are derived, as a function of m_A in the MSSM, by applying the analysis methods described in Chapter 7 to the data. Special care is taken to account for the increased natural width of the neutral Higgs bosons in the MSSM at high $\tan\beta$. The results are then extrapolated to higher luminosity data sets that should be available in the near future. A brief study is made of the quadruple b-tagged events, which could be used to add to the signal sensitivity of the analysis in the future. Finally, the conclusions are summarized.

8.1   Results

The normalization of the expected background is based on a Gaussian fit to the shape of the expected signal's invariant mass distribution, derived from bh Monte Carlo. Table 8.1 shows the means and widths of the fits to the MC signal invariant mass distributions, after applying the optimized sets of cuts. The means are roughly identified with the Higgs mass in each case. The width nearly doubles as m_h is increased from 100 to 120 GeV/c^2, mostly due to the reduction of turn–on effects from the jet E_T cuts. The width also rises simply because it is naturally a constant fraction of the Higgs mass, since the jet energy resolution is roughly proportional to the jet energy. The widths are roughly independent of whether events are used with ≥ 3 or ≥ 4 jets. The width is slightly smaller in the 3–jet case, where the jets from the Higgs decay can be identified more accurately. The events with fewer jets pose less combinatoric difficulty.

Table 8.1: Means and widths of the Gaussian distributions fitted to signal, after optimized analysis cuts, for each m_h, in the 3 and 4 jet cases.

Signal (GeV/c^2)	Mean (GeV)	Width (GeV)
n_j^{min}=4, m_h=100	104	12
n_j^{min}=4, m_h=120	123	23
n_j^{min}=4, m_h=150	136	26
n_j^{min}=3, m_h=100	103	14
n_j^{min}=3, m_h=120	121	19
n_j^{min}=3, m_h=150	143	25

The cuts made in this analysis are made in three stages: the trigger level, the kinematic analysis cuts (E_T, η, n_j, etc.), and b–tagging cuts. Table 8.2 shows the acceptance of each set of analysis cuts, for each m_h. The trigger efficiency rises slightly as a function of the mass of the Higgs, as expected. It is also slightly higher for the 4–jet events than for the 3–jet, since there is frequently an extra jet to trigger on. The kinematic cuts are also satisfied more often for a heavier Higgs boson, for similar reasons. The b–tagging efficiency is roughly independent of the Higgs mass or the number of jets in the events [1].

8.1.1 Results Without Higgs Boson Widths

The natural widths of the neutral Higgs bosons are dependent on $\tan\beta$. They are roughly proportional to $\tan^2\beta$ for the observable neutral Higgs bosons at high $\tan\beta$. Limits on $\tan\beta$ are first set for each neutral Higgs boson mass, ignoring the widths. The widths of the Higgs bosons at these excluded $\tan\beta$ limits are then taken into account, as described in the next Section.

Table 8.3 shows the number of weighted background events expected (with errors),

---

[1]The slight decrease in b–tagging efficiency at higher m_h is due to a slight decrease in b–tagging efficiency for high-E_T jets (E_T>75 GeV).

Table 8.2: The acceptance for signal of each set of analysis cuts, for each m_h, in the 3 and 4 jet cases.

Signal (GeV/c^2)	Kinematic (%)	Trigger (%)	B-tagging (%)	Total (%)
n_j^{min}=4, m_h=100	1.25	69	23	0.20
n_j^{min}=4, m_h=120	2.40	71	21	0.36
n_j^{min}=4, m_h=150	3.92	79	20	0.63
n_j^{min}=3, m_h=100	4.75	59	17	0.48
n_j^{min}=3, m_h=120	8.55	62	17	0.89
n_j^{min}=3, m_h=150	13.2	71	16	1.5

the total signal acceptance (with errors), the number of weighted events seen in data, and the 95% C.L. limits derived, for each m_h studied.

Figure 8.1 shows the 95% C.L. limit on $\tan\beta$ set as a function of m_A, using the MCLimit analysis method. Allowing three or more jets has more sensitivity than only allowing events with four or more jets, for all Higgs masses studied. The sensitivity decreases as m_A rises, due to the falling cross–sections for production.

8.1.2 Results With Finite Higgs Boson Widths

Given that this analysis, with the current data available, is only able to exclude Higgs bosons at very high $\tan\beta$, the larger widths of the neutral Higgs bosons must be accounted for at these points in MSSM parameter space. For the $\tan\beta$ values excluded at 95% C.L. in the 3–jet case, the widths of the Higgs, as given by HDECAY [10], are shown in Table 8.4. For the largest m_h studied, the Higgs width is comparable to the detector di–jet mass resolution.

The increased width of the Higgs is simulated by smearing the Higgs resonance in each event, systematically changing the E_T of the jets which make up the Higgs mass peak by a Gaussian function with a width set to that of the Higgs. The effect of the Higgs width on each mass peak can be seen in Figure 8.2, for $m_h = 100$, 120,

Table 8.3: The number of Gaussian weighted background events expected, the total signal acceptance after Gaussian weighting (in percent), the number of Gaussian weighted events seen in data, the 95% C.L. limit on $\tan\beta$ for each m_h studied using the Gaussian fitting and weighting method (Limit 1) and using the MCLimit signal shape method (Limit 2), using the optimized set of cuts in the 3 and 4 jet cases (not including the effects of the natural width of the neutral Higgs bosons at high $\tan\beta$).

Signal (GeV/c^2)	Background	Acceptance	Observed	Limit 1	Limit 2
n_j^{min}=4, m_h=100	177±23	0.14±.04	189	105	110
n_j^{min}=4, m_h=120	192±22	0.31±.08	204	123	122
n_j^{min}=4, m_h=150	225±22	0.54±.13	262	181	166
n_j^{min}=3, m_h=100	339±24	0.23±.05	351	94	87
n_j^{min}=3, m_h=120	381±39	0.52±.12	396	103	93
n_j^{min}=3, m_h=150	449±44	0.92±.20	508	160	143
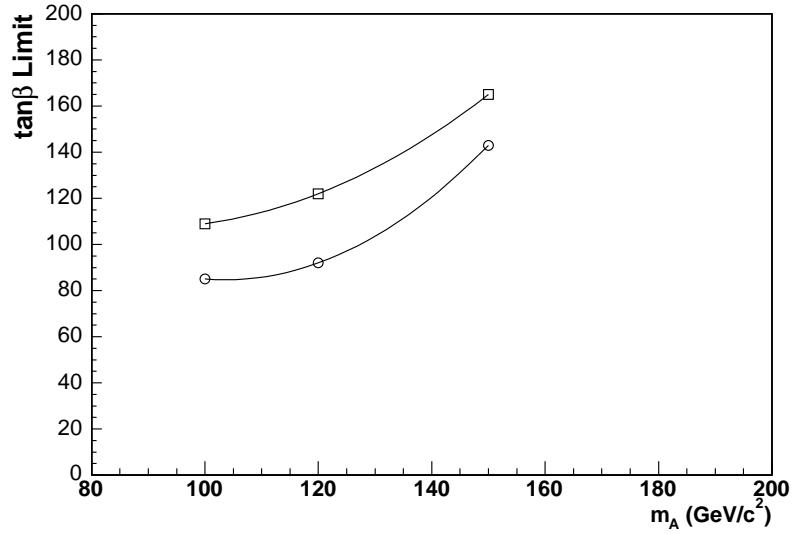
Figure 8.1: The 95% C.L. limit on $\tan\beta$ set as a function of m_A using the MCLimit analysis method (Limit 2, from Table 8.3), for the 3–jet cuts (open circles) and the 4–jet cuts (open squares). (The effects of the natural width of the neutral Higgs bosons at high $\tan\beta$ are not included.)

and 150 GeV/c^2. The peaks are fit with a sum of two Gaussian distributions.

There are two ways that the increased width of the Higgs can decrease the limit setting sensitivity. The first is that some signal efficiency is lost from the Higgs fluctuating low in mass and decaying to jets which do not pass the kinematic analysis cuts or may not have met trigger requirements. This effect is small because the Higgs boson is widest when its mass is largest and thus already past the kinematic and trigger turn–ons. A comparison of the number of signal events expected to survive all criteria for the final triple b–tagged data sample before and after the addition of the Higgs width is shown in Figure 8.2. Greater than 90% of all events are still accepted. A more detailed Pythia study [42] shows that >97% of signal is retained, after accounting for a proper re–weighting of the full event kinematics, as opposed to simple smearing at the jet level. Thus, the decrease in signal acceptance can safely be ignored.

The second way the increased Higgs mass can affect the expected limit is from the increased background included under the broader Higgs mass peak. However, a positive affect is that sometimes the Higgs will fluctuate high in mass into a region where there is less background.

The analysis cuts are re–optimized, using the smeared Higgs width signal distributions. The only change in the optimal analysis cuts is the leading jet E_T cut for the $m_h = 150$ GeV/c^2 case, which decreases from 60 GeV to 50 GeV. This is sensible since only the $m_h = 150$ GeV/c^2 case develops a large width relative to the detector resolution. The change in sensitivity between the new and old optimal leading jet E_T analysis cut was only 1.2% however, which does not significantly affect the resulting limit.

Both methods of limit setting, the Gaussian weighting method and the MCLimit method, are re-run with the added Higgs widths. The new limits obtained are shown in Table 8.4, and the results from the MCLimit method are shown in Figure 8.4. The limits are slightly weaker for the $m_h = 100$ and 120 GeV/c^2 cases, but are nearly

Table 8.4: The width of the Higgs at the $\tan\beta$ value excluded at 95% C.L., and the new $\tan\beta$ limits set after taking this width into account, using the two limit–setting methods (Gaussian is Limit 1 and signal shape is Limit 2).

Signal (GeV/c^2)	Higgs Width (GeV)	New Limit 1	New Limit 2
n_j^{min}=3, m_h=100	10	115	94
n_j^{min}=3, m_h=120	15	124	109
n_j^{min}=3, m_h=150	45	211	140

unchanged for the $m_h = 150$ GeV/c^2 case. Figure 8.3 shows the data, background, and Higgs mass peaks at the exclusion limit (where $\tan\beta = $ Limit 2 from Table 8.4), for the three Higgs boson masses studied.

8.2 Discussion

8.2.1 Quadruple b–Tagged Data to Monte Carlo Comparison

Although there are not yet enough events to set limits using the quadruple b–tagged data, it is instructive to compare it to Monte Carlo and the expected background distribution. These events may be used in future analyses to increase the overall signal sensitivity. The comparisons of the quadruple b–tagged data are shown in Figure 8.5. The normalization of the HF multi–jet backgrounds remains near unity (1.7±1.2), within statistical error. The Monte Carlo shape roughly describes the data.

8.2.2 Extrapolation to Higher Integrated Luminosity

Assuming that no deviation is observed between the data and the expected background in the future, the 95% C.L. limits on MSSM parameters which could be set given additional integrated luminosity can be predicted and are shown in Figure 8.6. The uncertainties on acceptance and background are assumed not to decrease as additional luminosity is acquired (which is conservative). The extrapolation also assumes
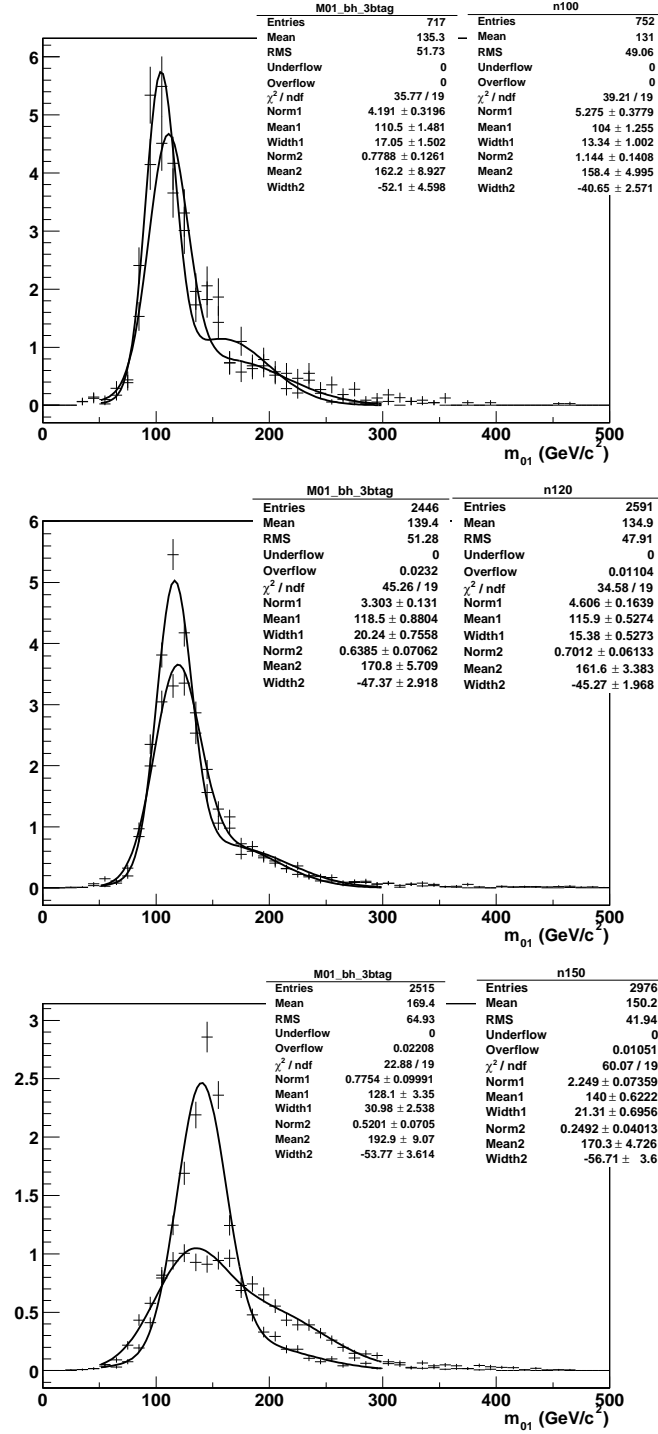
Figure 8.2: Comparisons between the invariant mass of the leading two jets in the triple b–tagged bh signal Monte Carlo before and after the addition of the Higgs width, for $m_h = 100$ (top), 120 (middle), and 150 GeV/c^2 (bottom).
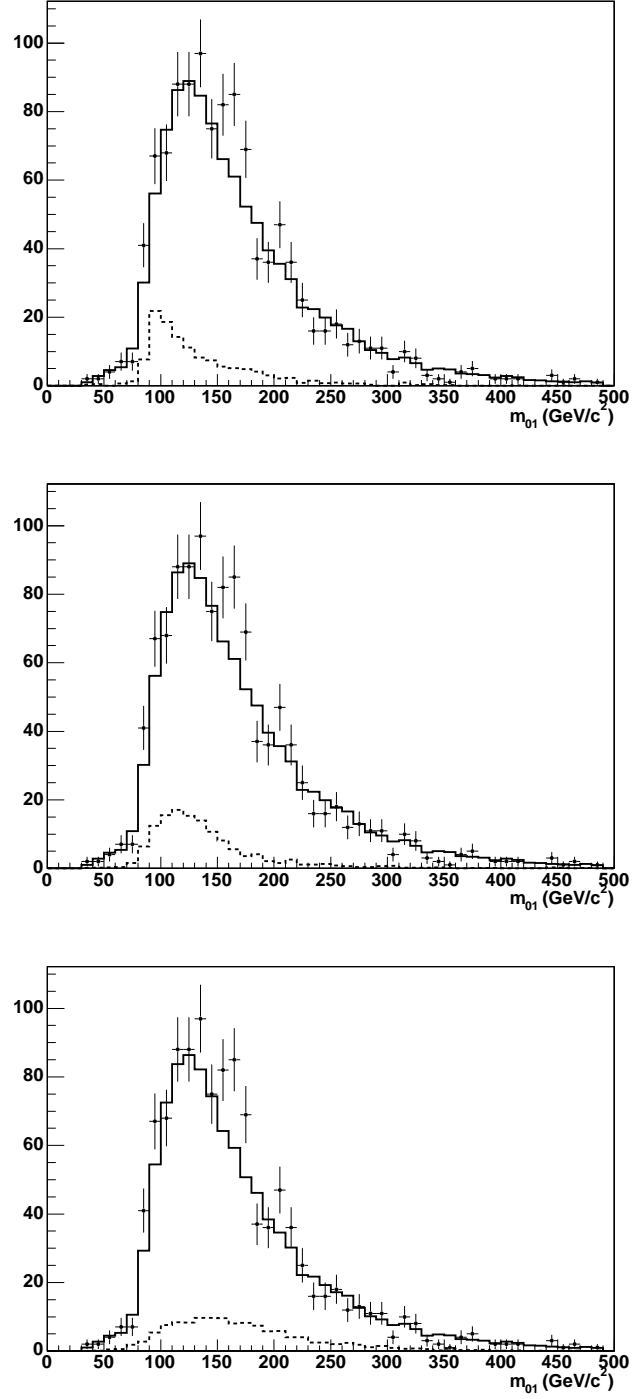
Figure 8.3: The data (points), background (solid), and Higgs signals (dashed) at the exclusion limit in the 3–jet case with 3 b–tags after the addition of the Higgs width, for $m_h = 100$ (top), 120 (middle), and 150 GeV/c^2 (bottom).
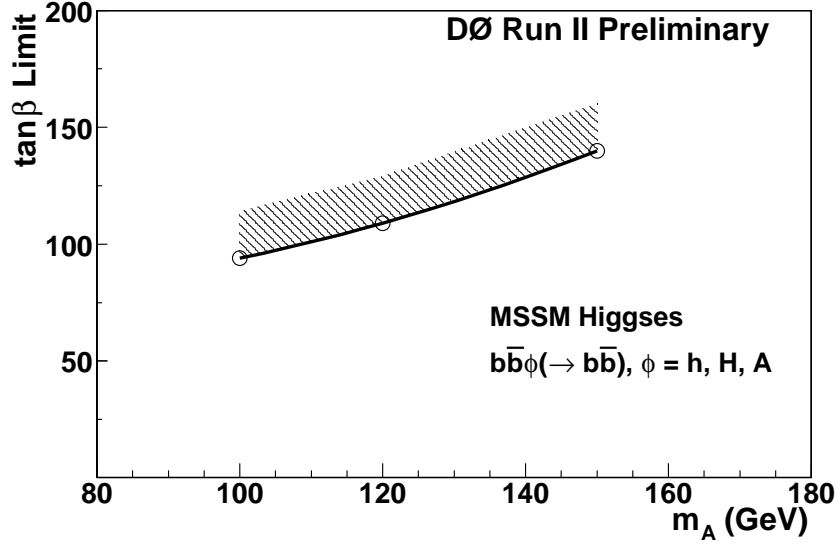
Figure 8.4: The 95% C.L. limit on $\tan\beta$ set as a function of m_A, which includes the effects of the Higgs widths, (Limit 2, from Table 8.4), for the 3–jet cuts. The hatched area indicates the direction in parameter space which is excluded.
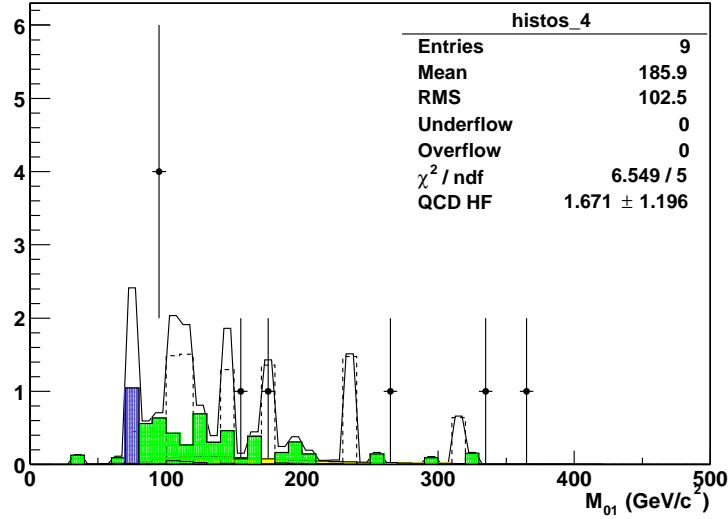
Figure 8.5: Fit of the quadruple b–tagged data's two leading jet invariant mass spectrum, after fake–tag parameterization correction, to a sum of backgrounds: multi–jet fakes (solid), ALPGEN $b\bar{b}$jj MC (dashed), and other small backgrounds ($Z(b\bar{b})$+jets (blue), $t\bar{t}$ (red), and ALPGEN $b\bar{b}b\bar{b}$ (green, wavy).
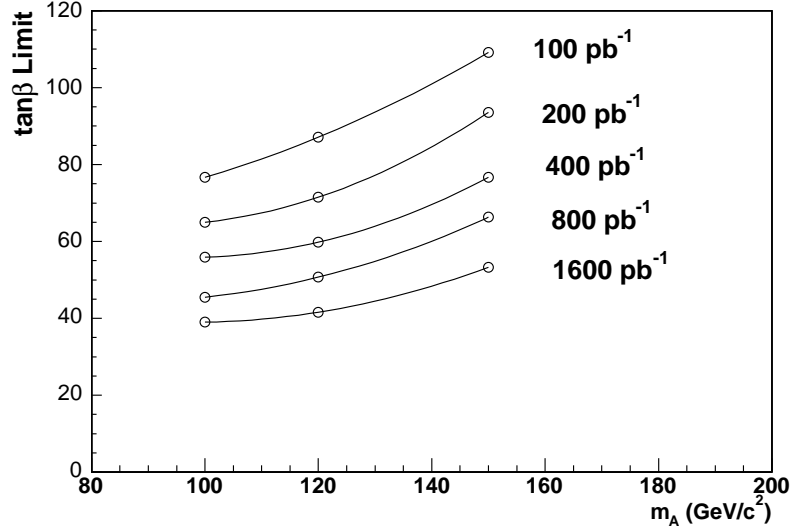
Figure 8.6: Extrapolation of the expected 95% C.L. upper limits on $\tan\beta$ in MSSM parameter space in the 3–jet channel, given perfect agreement between data and background with various integrated luminosities collected.

the current b–tagging performance, even though preliminary studies indicate significant improvements. The sensitivity can be greatly enhanced by decreasing the fake b–tagging rate (while keeping the b–tagging efficiency for real b–jets high), since the dominant backgrounds involve fake b–tagged jets. The widths of the neutral Higgs bosons are small compared to the detector resolution at the values of $\tan\beta$ which could be excluded with more data. The effects of the larger width of the Higgs bosons at high $\tan\beta$ are thus ignored during the extrapolation.

8.3   Conclusions

A search has been performed for a signal from neutral Higgs bosons in multi–jet events collected with the DØ detector during Run II of the Fermilab Tevatron. The results are interpreted in the context of the MSSM to provide limits in the m_A vs. $\tan\beta$ parameter space. For the m_A range studied ($100 - 150$ GeV/c^2), values of $\tan\beta > 80$

– 120 are excluded at 95% Confidence Level, as shown in Figure 8.4. Improvements in analysis techniques and larger data sets should allow significantly lower $\tan\beta$ values ($\gtrsim 50$) to be excluded in the near future.

BIBLIOGRAPHY

[1] S. Willenbrock, "Hadron Colliders, the Standard Model, and Beyond", hep-ph/0212032.

[2] LEP Higgs Working Group (July 2001), Note/2001-04, hep-ex/0107030.

[3] The CDF Collaboration, Phys.Rev.Lett. 86 (2001) 4472-4478, hep-ex/0010052.

[4] A. Haas, A. Kharchilava, and G. Watts, "Comparison of the DØ Search for Neutral Higgs Bosons at High $\tan\beta$ in Multi–jet Events to the CDF Run I Analysis", DØNote 4365.

[5] The Higgs Working Group: Summary Report (2001), hep-ph/0203056, pages 10-28.

[6] A. Nelson, L. Randall, "Naturally Large $\tan\beta$", hep-ph/9308277.

[7] V. Barger, C. Kao, "Implications of New CMB Data for Neutralino Dark Matter", hep-ph/0106189.

[8] D.I. Kazakov, "Beyond the Standard Model", hep-ph/0012288.

[9] M. Spira, "Higgs Boson Production and Decay at the Tevatron", hep-ph/9810289.

[10] M. Spira, "HDECAY", hep-ph/9704448, http://people.web.psi.ch/spira/hdecay/.

[11] J. Stark, http://www-clued0.fnal.gov/~stark/cal_corr_on_tmb.txt, October, 2003.

146

[12] "MC Runjob Documentation", http://www-clued0.fnal.gov/runjob/.

[13] A. Belyaev, E. Boos, "Single top quark tW+X production at the LHC: a closer look", hep-ph/0003260, Phys.Rev. D63 (2001) 034012.

[14] Private meeting with Steve Mrenna, summarized here:

http://www-d0.fnal.gov/Run2Physics/higgs/d0_private/groups/hbb/

talks/meet_1.html.

[15] J. Campbell, R.K. Ellis, F. Maltoni, S. Willenbrock, "Higgs-Boson Production in Association with a Single Bottom Quark", Phys.Rev.D67 (2003) 095002, hep-ph/0204093 v2.

[16] S. Dittmaier, M. Kramer, M. Spira, "Higgs Radiation off Bottom Quarks at the Tevatron and the LHC", hep-ph/0309204.

[17] S. Dawson, C.B. Jackson, L. Reina, D. Wackeroth, "Exclusive Higgs Boson Production with Bottom Quarks at Hadron Colliders", hep-ph/0311067.

[18] F. Maltoni, Z. Sullivan, S. Willenbrock, "Higgs-Boson Production vi Bottom-Quark Fusion", hep-ph/0301033.

[19] T. Sjstrand, P. Edn, C. Friberg, L. Lnnblad, G. Miu, S. Mrenna and E. Norrbin, Computer Phys. Commun. 135 (2001) 238 (LU TP 00-30), hep-ph/0010017.

[20] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions", JHEP 0307:001,2003, hep-ph/0206293.

[21] F. Maltoni, T. Stelzer, "MADEVENT: Automatic Event Generation with MAD-GRAPH", JHEP 0302:027,2003, hep-ph/0208156.

[22] E. Nurse and P. Telford, "Measurement of $\sigma \times$BR for Z$\rightarrow \mu^+\mu^-$ in $p\overline{p}$ Collisions at $\sqrt{s} = 1.96$ TeV", DØNote 4231.

[23] J. Campbell, R. K. Ellis, F. Maltoni, S. Willenbrock, "Associated Production of a Z Boson and a Single Heavy-Quark Jet", hep-ph/0312024.

[24] G. Blazey, et. al., "Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop", hep-ex/0005012.

[25] G. Blazey, et. al., "Run II Jet Physics", DØNote 3750, April 2002.

[26] S. Ellis, Private Communication, April 30, 2004.

[27] A. Goussiou, I. Iashvili, "Jet Energy Scale and Resolution for p13 Data and Monte Carlo",

http://www-d0.fnal.gov/phys_id/jes/d0_private/certified/v4.2/note.ps.

[28] L. Duflot, J.F. Grivaz, M. Ridel, P. Verdier, "Jets in High E_T events: data vs. Monte Carlo", DØNote 4020.

[29] JetMet group (conveners: G. Bernardi and V. Zutshi), "Jets in Run II", DØNote 3985.

[30] A. Schwartzman, M. Narain, "B-quark Jet Identification via Secondary Vertex Reconstruction", DØNote 4080.

[31] L. Feligioni, A. Schwartzman, L. Sonnenschein, M. Narain, "Update on the B-quark Jet Identification with Secondary Vertex Reconstruction using DØReco version p13", DØNote 4182.

[32] D. Bloch, B. Clément, D. Gelé, S. Greder, Isabelle Ripp-Baudot, "Performance of the Jet LIfetime Probability Method for B-tagging in DØ Data", DØNote 4158.

[33] Tobias Golling, "Presentation at the April 2003 b-id Meeting", slides are on the agenda server.

[34] The CDF Collaboration, Phys. Rev. D64(2001) 032002-1.

[35] B. Clément, D. Bloch, D. Gelé, S. Greder, A.C. Le Bihan, I. Ripp-Baudot, "SystemD or how to get signal, backgrounds and their efficiencies with real data", DØNote 4159.

[36] A. Goussiou, I. Iashvili, "Jet Energy Scale and Resolution for p13 Data and Monte Carlo",

http://www-d0.fnal.gov/phys_id/jes/d0_private/certified/v4.2/note.ps,

pages 94–125.

[37] A. Haas, "Level 1 and Level 2 Calorimeter Trigger Performance in Multi-jet Events", DØNote 4073.

[38] I. Bertram, G.Landsberg, J. Linnemann, R.Partridge, M. Paterno, H.B. Prosper, "A Recipe for the Construction of Confidence Limits", DØNote 3476.

[39] T. Junk, "Confidence Level Computation for Combining Searches with Small Statistics", Nucl.Instrum.Meth. A434 (1999) 435-443, hep-ex/9902006.

[40] The DØ Luminosity Group, "Luminosity Measurement", DØNote 4328.

[41] J. Rani, "Jet Efficiency Studies in γ+jet Events: Data vs. Monte Carlo", DØNote 4363.

[42] A. Kharchilava, "Pythia Higgs Width Study",

http://www-d0.fnal.gov/Run2Physics/higgs/d0_private/groups/hbb/

d0note4290comments/higgs_width.txt

[43] www.vmic.com

[44] www.tundra.com

[45] www.bvmltd.co.uk

[46] D. C. Schmidt and S. D. Huston, www.cs.wustl.edu/~schmidt/ACE.html, "C++ Network Programming, Vol. 1: Mastering Complexity with ACE and Patterns", AddisonWesley, 2001.

[47] xml.apache.org/xerces-c/

[48] fermitools.fnal.gov/abstracts/rgang/abstract.html

[49] www.linuxhq.com/kernel/v2.0/doc/networking/so_bindtodevice.txt.html

[50] www.gzip.org/zlib/

[51] www.oberhumer.com/opensource/lzo/

Appendix A

LEVEL 1 AND LEVEL 2 MULTI-JET TRIGGER MODELING

A.1 Introduction

Multi-jet signals, having two or more jets as distinguishing features, are a very important class of physics signals at DØ in Run II. For instance, neutral Higgs Bosons, with enhanced coupling to the b-quark (as predicted in Supersymmetry at high $\tan \beta$), could be observed as a three or four jet signal. Top-quark production at the Tevatron also has multi-jet signals and backgrounds. Understanding quantitatively the efficiencies with which such physics processes can pass the required trigger terms is needed for measuring (or placing limits on) their cross-sections.

The DØ trigger for Run II is composed of three levels; the first two levels are considered here. Multi-jet trigger terms at Level 1 make use of the L1 Calorimeter Trigger, which makes available the total transverse energy (E_T) of the 4 most energetic 'trigger towers', 0.2 ϕ by 0.2 η regions of the calorimeter. A typical multi-jet trigger term requires a given number of towers, X, to each above a certain E_T threshold, Y. This trigger term is referred to as $CJT(X,Y)$. X is an integer from one to four, and the value of Y can be any of the four values contained in the current *reference set*, which are programmed into the L1 firmware. Other, more complicated terms are also possible, such as the AND of requirements, CJT(3,3)CJT(2,5). Taking the OR of individual terms may also be supported in the future, and could offer significant gains in efficiency. At Level 2, towers above threshold are used as seeds, and the energy in adjacent towers in a 5 by 5 grid is summed. No merging of these L2 'jets'

can take place, due to time constraints. A complete L2 term requires X jets above Y GeV in E_T, and is denoted L2J(X,Y). These terms could be AND-ed or OR-ed with other terms. Another common L2 trigger term is to require that the scalar sum of the E_T of L2 jets above a threshold be greater than a given value, Z, referred to as L2H_t(Z).

The ability of multi-jet events to satisfy trigger terms is complicated. First, the effects of single, isolated offline jets on the calorimeter trigger towers are studied. Then, methods for combining the effects of all the jets in each event are presented, both for Level 1 and for Level 2. The quality of the combination techniques is cross-checked on an independent data sample, simply by comparing the predicted event trigger efficiencies with the actual efficiencies. The multi-jet efficiencies in data are then compared with the standard Monte Carlo trigger simulations.

A.2 Data Sample

The data taken from August – October of 2002 (about $10pb^{-1}$) was reconstructed with the p11 version of the DØ reconstruction software. Of this sample, events were 'skimmed' which had passed a muon L1 trigger, in order to minimize the bias in studies of jet triggers. Then, only events with one or more reconstructed cone jets (of radius 0.5) with (uncorrected) $E_T > 10$ GeV were kept, since only these events are useful for studying jet triggers. This data set was divided into two independent data sets, the only relevant difference being the time interval during which the events were recorded. The first contains about 65,000 events and was used for modeling the trigger performance. The second sample contains a comparable number of events and is used for cross-checking the method's predictions.

A.2.1 Jet Selection Criteria

The certified jet quality cuts were applied to each offline jet in the samples. They were:

- 0.05 <EM Fraction <0.95 (the fraction of transverse energy in the 'Electro-Magnetic' layers (1–7) of the calorimeter)

- Coarse Hadronic Fraction (CHF) <0.4 (transverse energy fraction in the coarse hadronic layers of the calorimeter)

- Hot Fraction <10 (the ratio of transverse energy in the most energetic cell to the next most energetic cell)

- n90 >1 (n90 is the number of cells required to contain 90% of the jet energy)

- f90 >0.8 – 0.5 * CHF (f90 is n90/nitm, and nitm is the total number of towers in the jet) This removes fake jets due to noise in the coarse hadronic calorimeter.

In addition, events from runs with Calorimeter problems were excluded.

A.2.2 Data Quality Checks

Figure A.1 contains basic plots showing the quality of the jets in the sample. While there is a warm region in η-ϕ, there are no bumps in the E_T spectrum, and the jets in the warm region comprise less than 10% of jets.

A.3 Level 1

Whether or not a particular multi-jet event has enough L1 Calorimeter trigger towers above the required thresholds to pass the trigger requirement is a function of the properties of all the jets in the event. There are several ways, for instance, even
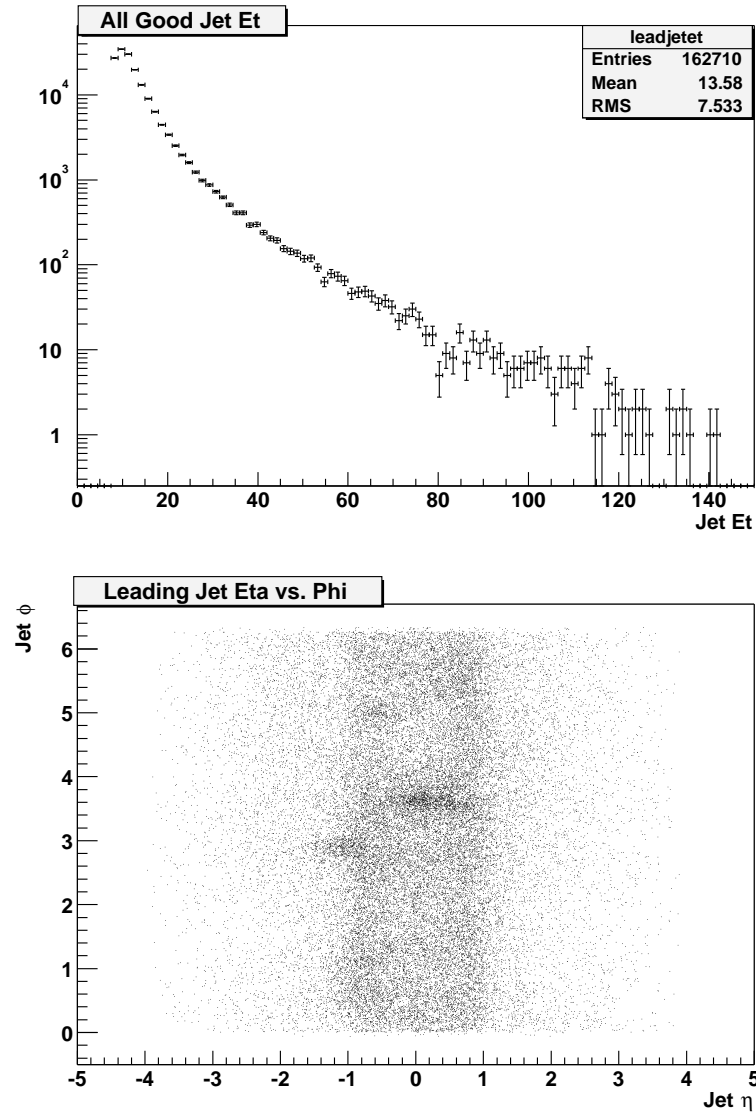
Figure A.1: The E_T distribution of jets passing quality cuts (top) and the η-ϕ distribution (bottom).

a simple di-jet event could deposit energy in two towers above a given threshold. Either each jet could fire one tower, or one of the jets could fire both towers. Also, the probability of a jet firing 0,1,2,... towers above threshold is a function of the jet's E_T, at least. The number of ways for 6 jets to fire CJT(3,3)CJT(2,5) is enormous!

First, the effects of single, isolated jets on L1 calorimeter trigger towers are measured. Specifically, a parameterization, by jet E_T, is derived of the average number of towers 'fired' (caused to be above threshold). Then, the effects of all the jets in a multi-jet event are combined by simply adding the expected average numbers of towers fired for each jet. This distribution of expected total number of towers fired is fit by a 'Scaled' Poisson function, and sampled to predict the probabilities for firing X or more towers above Y GeV. These requirements could then be used alone, or combined in AND or OR, with others to form a complete L1 calorimeter trigger.

The results of the predictions agree well with the measurements made on the independent data sample. A comparison of the data with Monte Carlo shows that the trigger is slightly less efficient in data than in the simulations.

A.3.1 Single-jet Studies

To begin, the effects of single jets are studied. Only those trigger towers near the jet are considered. Looking at the plot of the difference between the jet ϕ and the leading E_T tower ϕ, Figure A.2, a cut of 1.5 in $\Delta\phi$ should be wide enough to capture the leading, and even 2nd, 3rd, and 4th leading towers, but tight enough to eliminate towers effected by a possible second non-reconstructed jet in the event. (The physics of jet-production naturally biases the single-jet sample to contain a large number of events which were truly di-jet events, but where one of the jets was not reconstructed.)

A.3.2 Parameterizing Single-jet Trigger Tower Effects

Figure A.3 shows the number of towers above 5 GeV fired on average by a jet, as a function of the jet E_T. The low-E_T part of the curve nicely follows the Error-
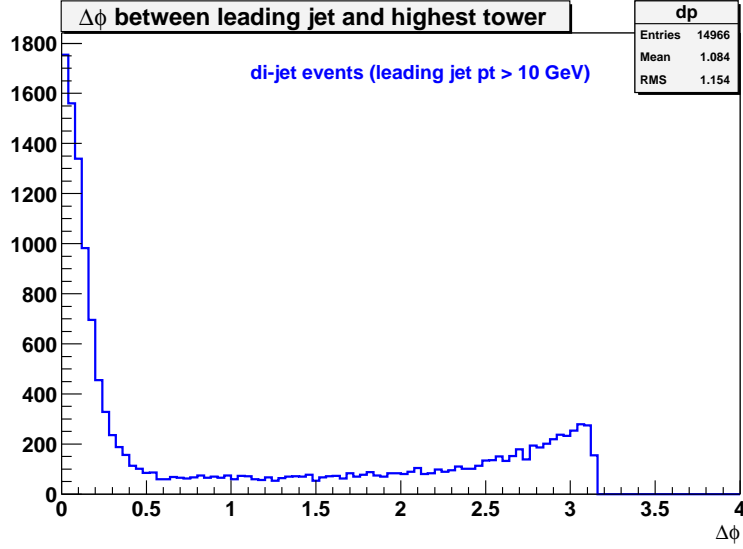
Figure A.2: The distribution of the difference in ϕ between the jet and leading E_T tower. A cut is applied to towers at 1.5 in $\Delta\phi$.

Function, or *erf* function (the integral of a Gaussian), because the dominant effect causing the slow 'turn-on' is the energy resolution of the towers. If the towers had perfect energy resolution, then the turn-on would be much steeper, although still not a perfect step function, since low E_T jets naturally distribute their energy amongst many towers. At high E_T, these geometric effects of energy distribution in towers dominates. The tendency of jets to become more collimated at higher E_T (in the kinematic region studied) seems to balance the effect that there is more energy to distribute, and the curve flattens out to a maximum of about 3 towers above 5 GeV being fired on average.

The fit is made using the erf function merged at higher E_T into the hyperbolic tangent function, *tanh*.

$$T(x) = 1, x > 35 \text{ and } (x/35)^2, 20 < x < 35 \text{ and } 0, x < 20$$
$$C(x) = P_0 + P_1 erf(P_2 x - P_3) + T(x)P_4 tanh(x/P_5 - P_6)$$

Note that only jets with $|\eta| < 1.7$ are considered for inclusion in this curve, since the

divided	
Entries	52
Mean	66.73
RMS	21.27
χ^2 / ndf	35.67 / 38
p0	0.49 ± 0.02946
p1	0.489 ± 0.03165
p2	0.1309 ± 0.007549
p3	2.66 ± 0.121
p4	2.188 ± 0.5493
p5	56.55 ± 19.25
p6	0.36 ± 0.1298

Figure A.3: The average number of towers above 5 GeV fired by a jet, as a function of the jet E_T (offline, no Jet Energy Scale corrections applied). The fit is done with the custom erf-tanh function (see text).

upper coverage of the L1 calorimeter towers is currently 2.4.

To be able to predict the efficiency of multi-jet events passing terms involving other thresholds, such as requiring 3 towers above 3 GeV, another parameterization would have to be made of the average number of towers caused to have an energy above 3 GeV by a single jet. For each 'reference set' of the calorimeter trigger, fortunately a limited number (4), a parameterization of the single jet effect is needed.

A.3.3   Predicting Multi-jet Efficiency

Now that the effects of single jets have been parameterized, the effects of the jets in a multi-jet event can be combined into a single prediction (a probability) for whether trigger terms such as CJT(X,Y) will pass. The jets are assumed to be isolated enough from each other so that they do not overlap, and thus their effects add linearly. The total number of towers expected to be above 5 GeV on average for a whole event is

Number of towers above 5 GeV in di-jet events with leading Et between 30 and 50 GeV

Gausian vs. Poisson Fits
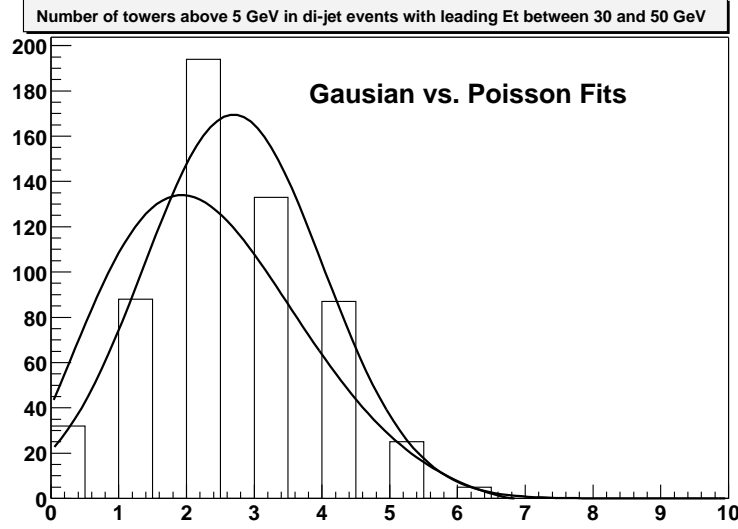
Figure A.4: The distribution of the number of towers above 5 GeV fired in di-jet events. Two fits are shown, to show that neither a Gaussian nor a Poisson correctly describes the distribution.

simply the sum of the average number of towers expected to be above 5 GeV for each jet in the event (with $|\eta| < 2.4$, i.e. in the fiducial L1 calorimeter coverage). In other words, for each jet in the fiducial calorimeter tower region, the expected number of towers fired is looked up on the erf-tanh parameterized single-jet effect curve above, and added to the total expected for the event.

Predicting the average number of towers above threshold is still one step away from the goal of predicting how often there will be X towers above 5 GeV. To predict the probability that X or more towers will be above 5 GeV, the distribution of numbers of towers above threshold needs to be studied, not just the average. The distribution of the number of towers above 5 GeV is shown for di-jet events in Figure A.4. Correctly describing this distribution is critical, since it is what will be used to predict the efficiency of trigger terms. Neither the Gaussian nor the Poisson distributions correctly describe the data. The 'correct' distribution seems to lie somewhere in between.
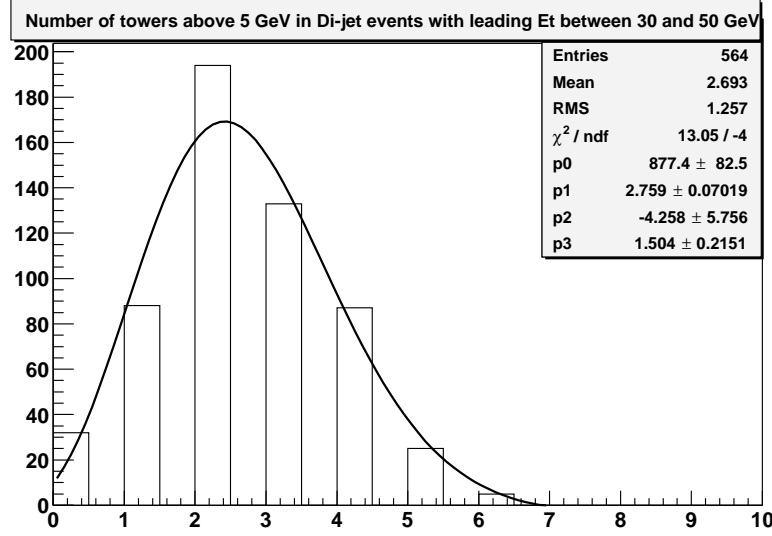
Figure A.5: The distribution of the number of towers above 5 GeV fired in di-jet events with leading jet E_T between 30 and 50 GeV. The Scaled Poisson describes the distribution better, with a scale factor of 1.5.

The Poisson distribution approaches the Gaussian distribution as the mean of the distribution increases. Therefore, a distribution that lies between the Poisson and Gaussian distributions is simple to create. The 'Scaled' Poisson distribution is a standard Poisson distribution, but with an extra parameter, the scale factor, α. This scale factor multiplies the mean of the distribution, and is used consistently to normalize the distribution so that its integral remains unity.

$$f(x) = \alpha e^{-\alpha\mu} \frac{(\alpha\mu)^{\alpha x}}{\Gamma(\alpha x+1)}$$

The Scaled Poisson distribution approaches a Gaussian α times faster than a standard Poisson. The data are now well described, as seen in Figure A.5.

Other jet energy ranges are also studied, and proper scale factors are chosen to describe them correctly. Figures A.13 and A.14 show the low and high E_T ranges, respectively. The scale factor which fits the data follows the trend:

$$\alpha = 1 + \left(\tfrac{\mu}{4}\right)^2$$

where μ is the (unscaled) mean of the distribution.

Once the distribution of expected total towers above threshold is known, then it is trivial to predict the probability that CJT(2,5), for instance, will fire. The chance of 2 or more towers firing (above 5 GeV) is 1-P(0)-P(1), where P(n) is the probability of exactly n towers firing. P(n) can be found immediately by consulting the Scaled Poisson distribution function just by replacing x with n.

A.3.4 Testing the Multi-jet Efficiency Predictions

There is now a complete prescription for predicting the efficiencies for L1 calorimeter terms in multi-jet events, based on the E_T's of all the jets in the event (within the fiducial L1 calorimeter tower region). For each event the predicted probability of passing a given trigger term, CJT(X,5), is determined. This prediction can be compared with the actual firing of that given trigger term for each event. The independent data sample (different from the sample used to derive the prediction curves) is used for the comparisons. Whereas all jet E_T's are used for calculating the predictions, the comparisons are plotted as a function of one variable only, the leading jet E_T, for visual clarity.

Figure A.6 shows the comparison of data and predicted CJT(3,5) turn-on. Plots are shown of other trigger terms in Figures A.15–A.17, as well as a comparison with non-Scaled Poisson fits to the data distribution in Figures A.18 and A.19. Clearly, the data is better described when the Scaled Poisson method is used to model the distribution of the number of trigger towers above threshold.

A.3.5 Comparison with Monte Carlo Trigger Simulations

The simulations of the trigger in Monte Carlo should agree closely with the data. If there are discrepancies, it is important to understand their magnitudes and be able
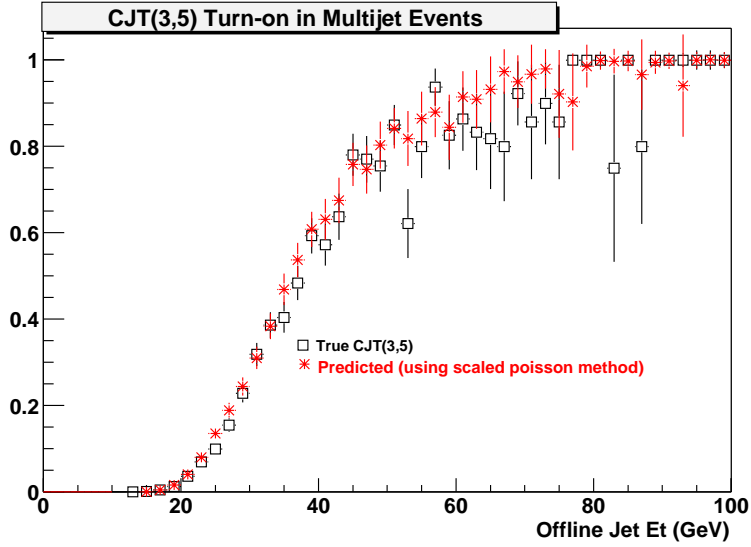
Figure A.6: Comparison of data and prediction using the Scaled Poisson method for CJT(3,5).

to correct for them. The rate of CJT(3,5) in 'QCD' events (generated with a parton p_T cut of 20 GeV) is seen to be a factor of 1.5 lower in data than predicted in Monte Carlo. The disagreement arises largely in the low-E_T part of the spectrum, as seen in Figure A.7. This could be due to Monte Carlo simulation effects of the parton p_T cut, or an overestimation (in the simulation) of the energy resolution of the towers. It is encouraging that this rate discrepancy is verified directly: online L1 calorimeter pass rates (for all terms, uniformly) are observed to be a factor of 1.6 times lower than would be expected for the measured luminosity, predicted QCD 20 GeV cross-section, and standard Monte Carlo trigger simulations.

Comparisons of data-based predictions and standard MC simulations of CJT(3,5) efficiency for bbh signals are shown in Figures A.20 and A.21.
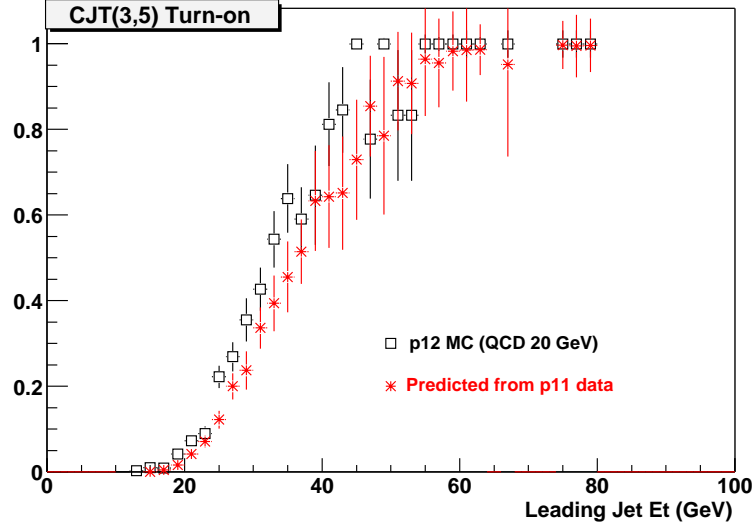
Figure A.7: Comparison of Monte Carlo predictions and the new data-based prediction using the Scaled Poisson method for CJT(3,5) in QCD events with a generator-level parton p_T cut of 20 GeV.

A.4 Level 2

Combining the effects of multiple jets at Level 2 is simpler than at Level 1, since to first order each offline jet only fires a single L2 jet, whereas at Level 1, the chance of a single jet firing 2, 3, or even 4 towers is equally important. However, it is necessary to include the rate at which single offline jets create two L2 jets above threshold to produce acceptably accurate predictions. The turn-on curve, as a function of offline jet E_T, for firing a L2 jet above threshold is measured for single jets. To combine the effects of the many jets in each multi-jet event, rigorous probability accounting is used to build up the distribution of chances for firing 0, 1, 2... jets above threshold, being careful to account for the chance that a second L2 jet above threshold may also result from a single offline jet.

The predicted multi-jet turn-on curves for requiring 1, 2, and 3 L2 jets above threshold are checked against the independent data set, and the agreement is sat-

isfactory, as shown in Figures A.23, A.24, and A.10. And Figure A.11 shows good agreement between the predicted efficiencies and rates derived in simulated Monte Carlo events.

Another trigger term used at L2 is the scalar sum of the E_T of jets above a threshold (normally 5 GeV), a quantity known as L2 H_t. The turn-on of L2 H_t is studied versus offline H_t, and parameterized. This measurement is then compared to Monte Carlo simulations of L2 H_t, and the H_t in data is seen to turn on more slowly than in MC.

A.4.1 Single-jet Efficiencies

Similar to the way single jet effects were studied for Level 1 towers, the effect of a single offline jet at Level 2 is now measured. Only Level 2 jets within a $\Delta\phi$ of 1.5 are considered to be associated with the offline jet, to protect against the effects of a possible second, non-reconstructed, p_T-balancing jet. The turn-on curve is measured here only for an 8 GeV L2 jet threshold, and would need to be re-measured for other L2 jet thresholds to study terms involving L2 jets of other E_T's.

The turn-on of L2 jets, shown in Figure A.8, is well described by a simple Error-function (erf). Only the Gaussian-smeared energy resolution of the L2 jets affects the shape of the turn-on. The fit function used is:

$$C(x) = P_0 + P_1 erf(P_2 x - P_3)$$

A.4.2 Contribution of 'Double' L2 Jets

One other 'second-order' effect must be accounted for, to accurately describe the data. About 10% of the time, a single offline jet creates more than one L2 jet above threshold. This effect is parameterized as a function of offline jet E_T, and shown in Figure A.9. The distribution of the number of L2 jets above threshold to be associated with the single offline jet is shown in Figure A.22.
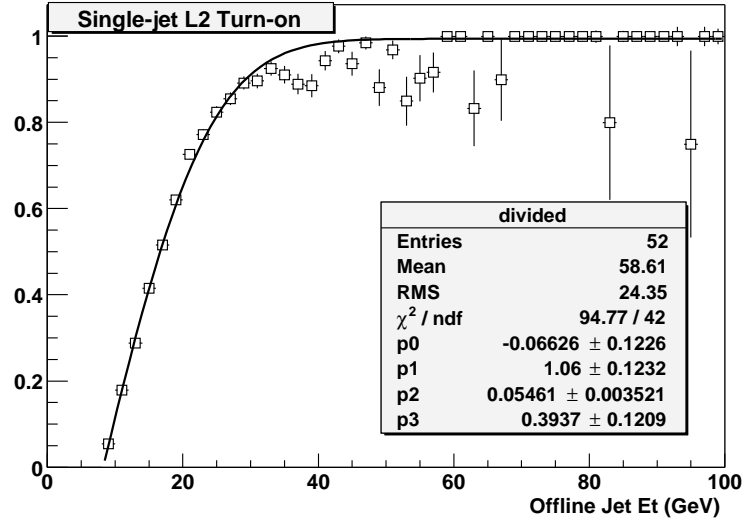
Figure A.8: Turn-on of L2 jet above 8 GeV (E_T) associated with offline jet in single-jet events.
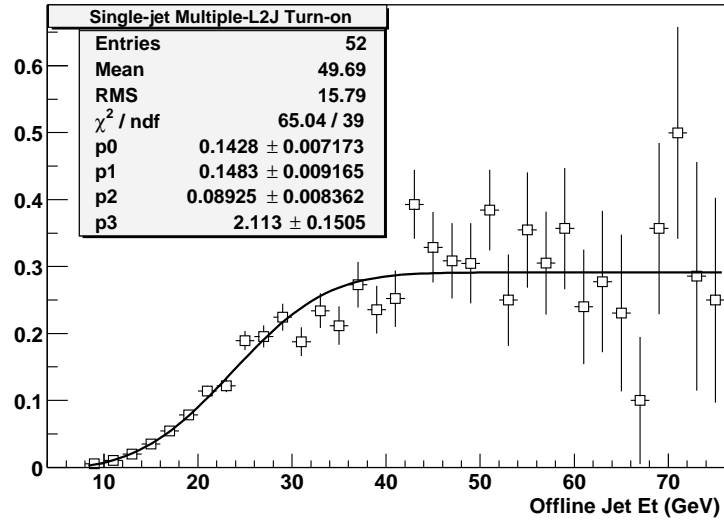
Figure A.9: The fraction of events in which a single jet is associated with more than one L2 jet, as a function of the offline jet E_T.

A.4.3   Combining into Event Efficiency

Knowing the probabilities for creating 0, 1, or 2 L2 jets above threshold, as a function of the jet's E_T (the probability of creating more than 2 L2 jets from a single jet is ignored since, as seen in Figure A.22, this occurs in <0.1% of jets), it is now straightforward to construct the probability distribution for X L2 jets to be above threshold for an entire multi-jet event. Consider just the probability of exactly 1 L2 jet being above threshold, PE(1). After the first offline jet (within $|\eta| < 2.4$, the fiducial region of calorimeter trigger towers), the probability of exactly 1 L2 jet firing for the event is $P(1,E_T{}^1)$, i.e. the probability of exactly 1 L2 jet firing above threshold, given the offline jet having E_T of $E_T{}^1$. After the second offline jet, the PE(1) is modified to PE(0)*$P(1,E_T{}^2)$ + PE(1)*$P(0,E_T{}^2)$, i.e. the new chance that exactly one L2 jet will be above threshold is the probability that the first jet didn't fire and the second one did, or the first jet did fire and the second one did not. This method can be applied to every successive offline jet. For instance, PE(1) is modified for the Nth jet as PE(0)*$P(1,E_T{}^N)$ + PE(1)*$P(0,E_T{}^N)$.

To account for the small probability that 2 L2 jets will be fired above threshold by a single offline jet, PE(2+n) is adjusted after each offline jet to be PE(1+n)*$P(2,E_T{}^N)$ + PE(2+n)*$P(0,E_T{}^N)$. In other words, it is dealt with exactly as if for each offline jet there was a second offline jet, but the second one had a much smaller probability of firing a L2 jet above threshold, $P(2,E_T{}^N)$, than the first, $P(1,E_T{}^N)$.

A.4.4   Testing the Multi-jet Efficiency Predictions

Using the complete calculated probability table for PE(X), it is simple to calculate the probability for X or more L2 jets to be above threshold. For instance, the probability for there to be 3 or more L2 jets above threshold is 1 - PE(0) - PE(1) - PE(2). The comparison of the predicted L2J(3,8) turn-on in multi–jet events is shown in Figure A.10. Although all jet E_T's are used to calculate probabilities, for clarity the
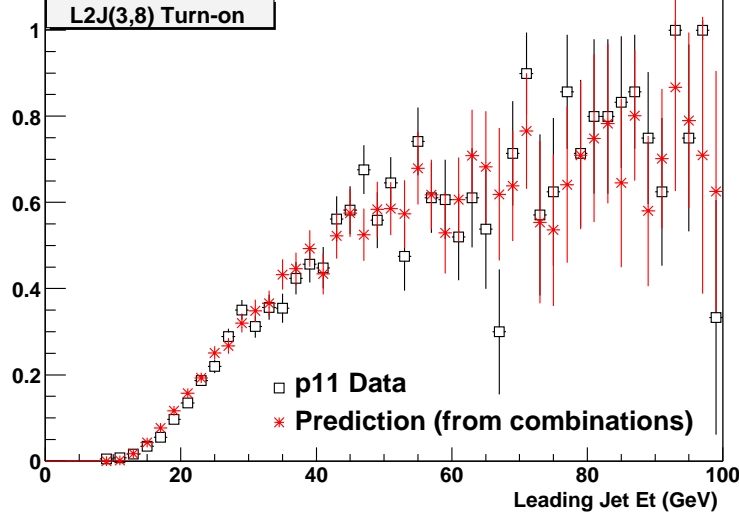
Figure A.10: Turn-on of L2J(3,8) in multi-jet events.

turn-on is shown as a function of the leading jet E_T only.

The turn-ons for other numbers of L2 jets are shown in Figures A.23 and A.24. Also, Figure A.25 shows the disagreement for L2J(3,8) between predictions and data when the chance of a single offline jet firing 2 L2 jets above threshold is not accounted for.

A.4.5 Comparison with Monte Carlo Trigger Simulations

A comparison of the data-based L2J(3,8) trigger prediction to the standard Monte Carlo trigger simulation is shown in Figure A.11 for bbh($\rightarrow b\bar{b}$) signal events, with a Higgs mass of 120 GeV. The agreement is very good.

A.4.6 Level 2 H_t

Another L2 trigger term frequently used for multi–jet signals is to require L2 H_t, the scalar sum of the E_T of all L2 Jets above some threshold (usually 5 GeV), to be above
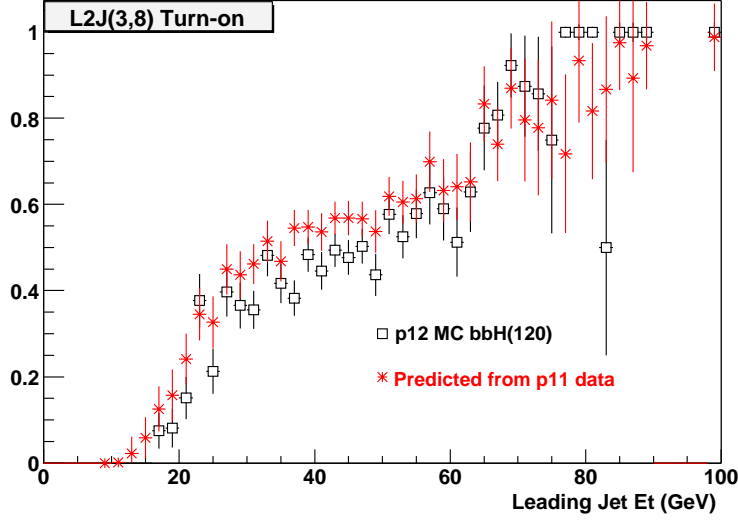
Figure A.11: Comparison of the turn-on of L2J(3,8) with the standard trigger simulation in bbh events, with m_h=120 GeV.

some threshold. Here the turn-on of L2H_t(50) is studied, which is used in the bbh trigger. First, the turn-on of L2H_t(50) is measured in multi-jet events, as a function of offline E_T. The plot is shown in the Appendix in Figure A.26. The curve is fit well with a simple Error-function, which is appropriate since only the Gaussian-smeared energy resolution of the towers effects the turn-on.

Figure A.12 shows the comparison between the standard trigger simulation and the data-measured turn-on curve, using bbh MC events as the sample (with a Higgs mass of 120 GeV).
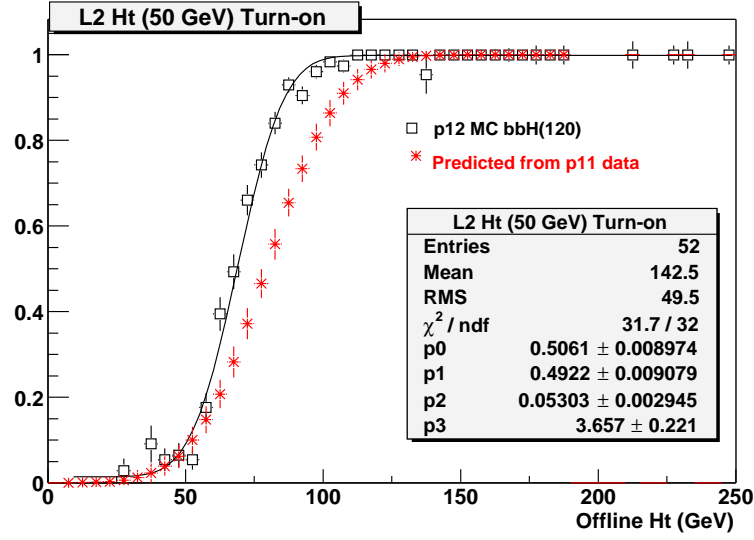
Figure A.12: Comparison of the turn-on of $L2H_t(50)$ in data with the standard trigger simulation in bbh signal events, with m_h=120 GeV.
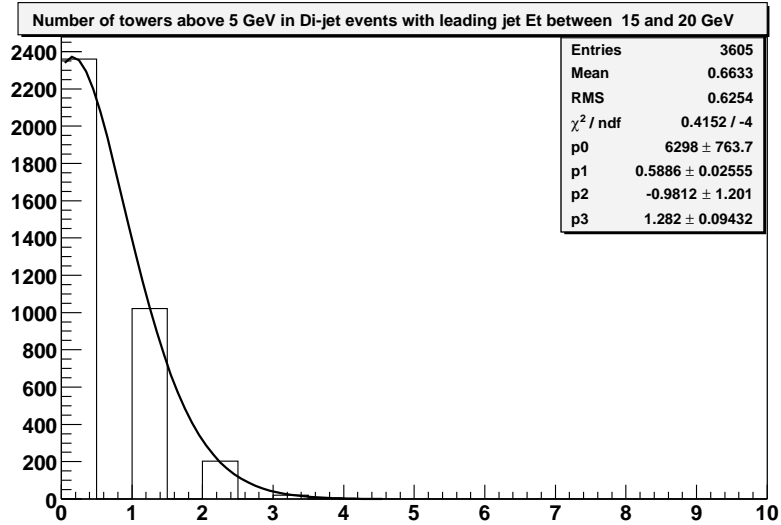
Figure A.13: The distribution of the number of towers above 5 GeV fired in di-jet events with leading jet E_T between 10 and 20 GeV. The Scaled Poisson correctly describes the distribution, with a scale factor of 1.3
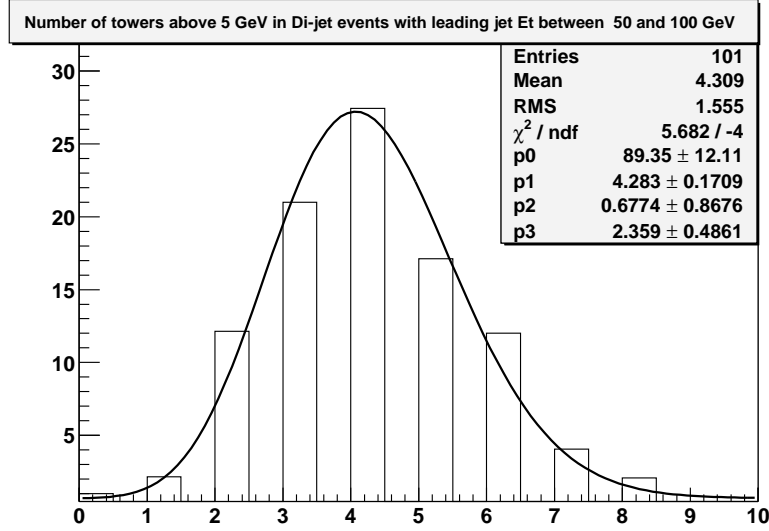
Figure A.14: The distribution of the number of towers above 5 GeV fired in di-jet events with leading jet E_T between 50 and 100 GeV. The Scaled Poisson correctly describes the distribution, with a scale factor of 2.4
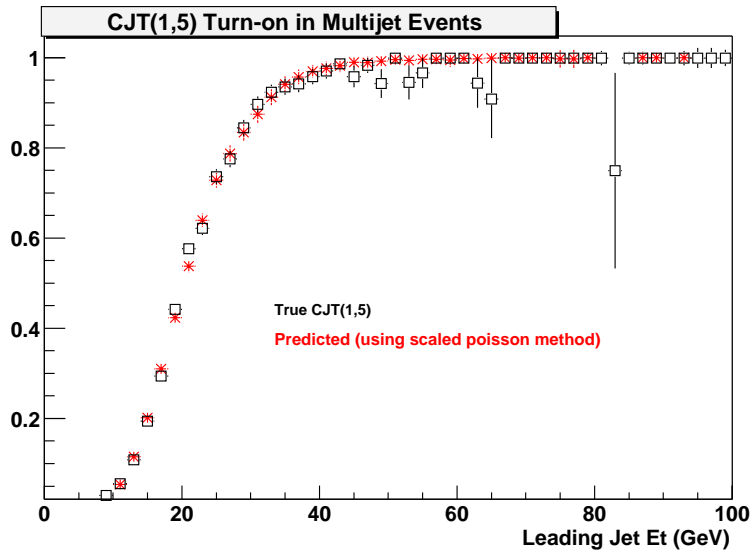
Figure A.15: Comparison of data and prediction using the Scaled Poisson method for CJT(1,5).
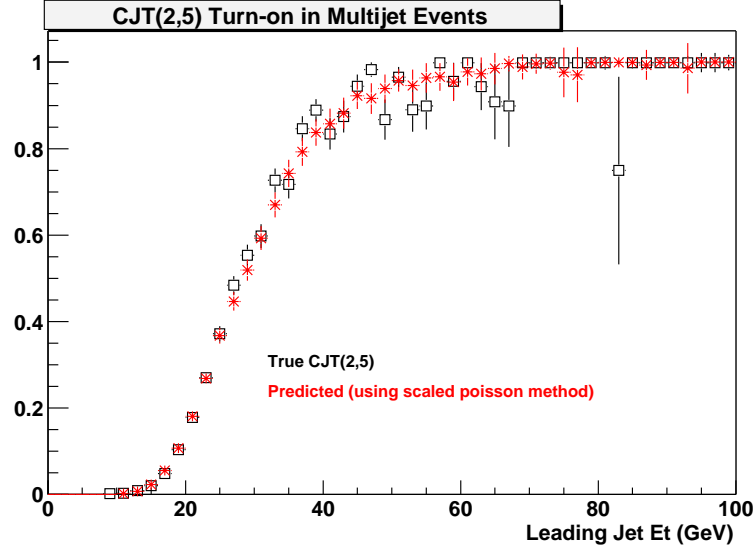
Figure A.16: Comparison of data and prediction using the Scaled Poisson method for CJT(2,5).
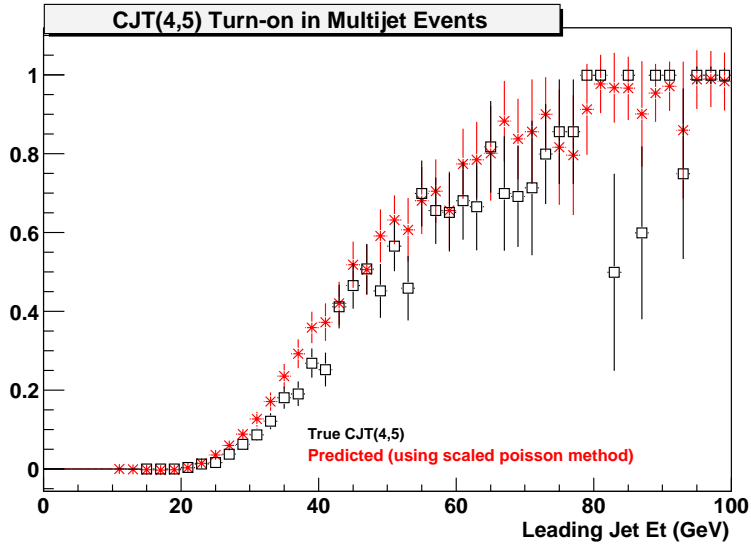
Figure A.17: Comparison of data and prediction using the Scaled Poisson method for CJT(4,5).
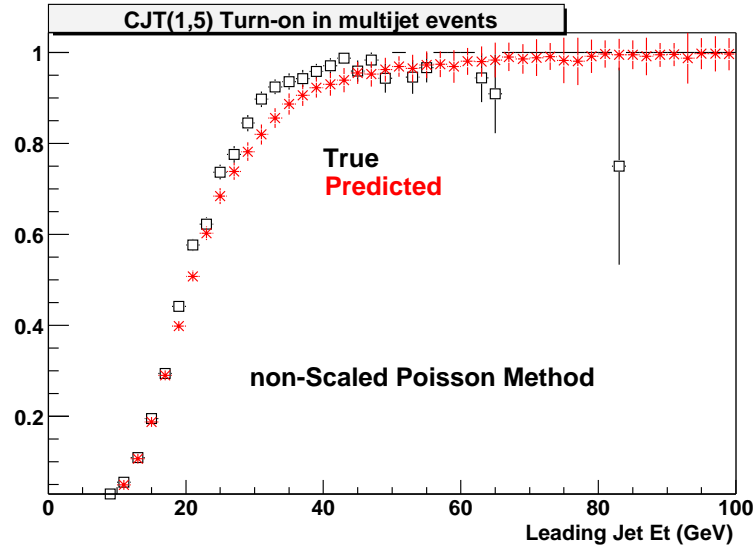
Figure A.18: Comparison of data and prediction using the Normal Poisson method for CJT(1,5).
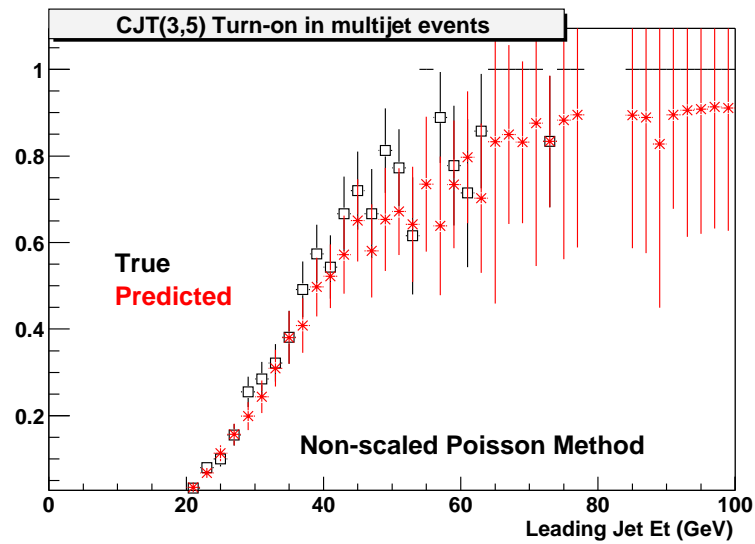
Figure A.19: Comparison of data and prediction using the Normal Poisson method for CJT(3,5).
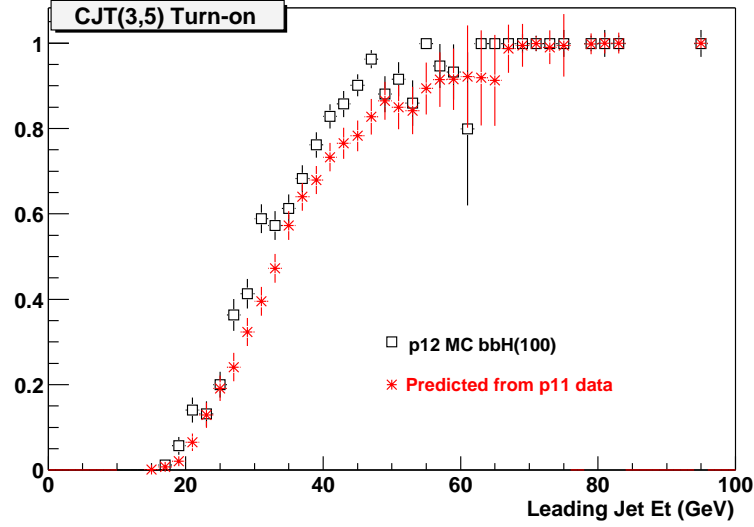
Figure A.20: Comparison of Monte Carlo predictions and the new data-based prediction using the Scaled Poisson method for CJT(3,5) in bbh($\rightarrow b\bar{b}$) events with a Higgs mass of 100 GeV.
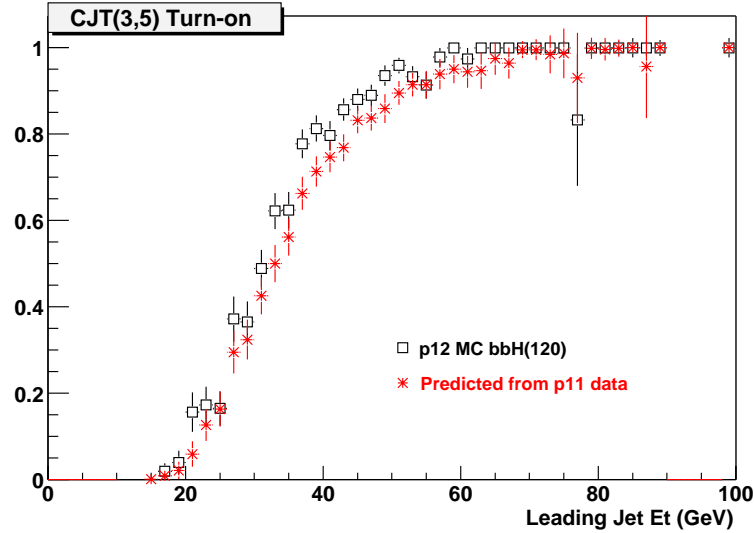
Figure A.21: Comparison of Monte Carlo predictions and the new data-based prediction using the Scaled Poisson method for CJT(3,5) in bbh($\rightarrow b\bar{b}$) events with a Higgs mass of 120 GeV.
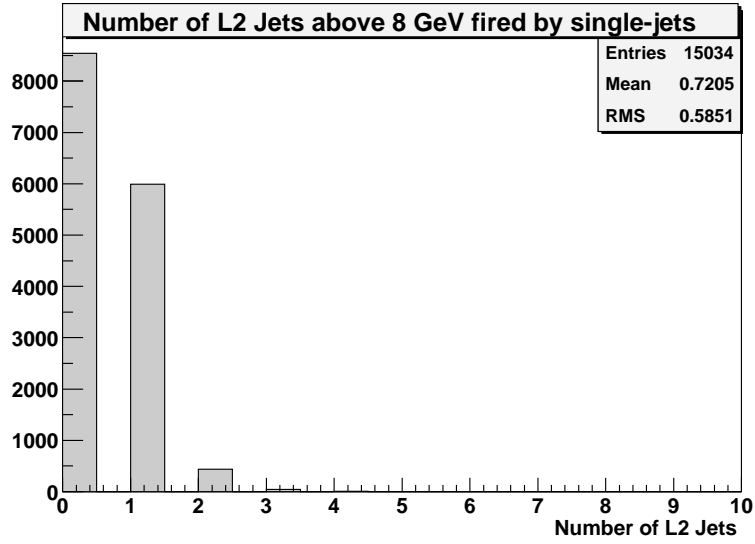
Figure A.22: The distribution of the number of L2 jets above threshold that are associated with a single offline jet.
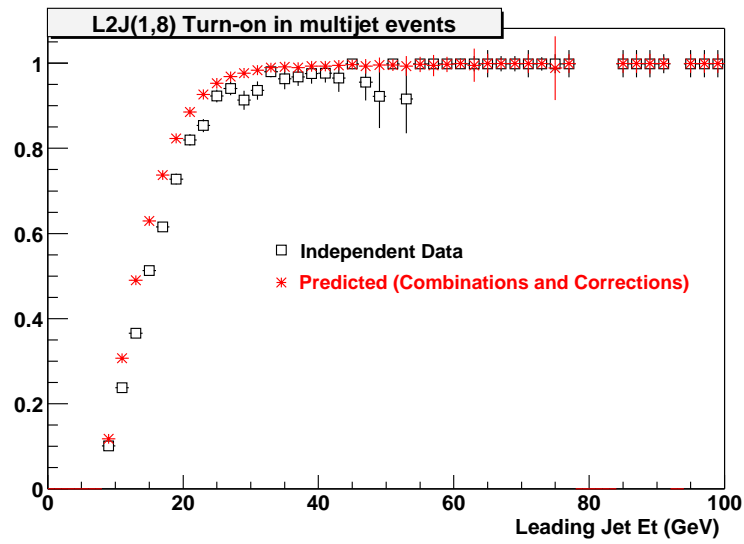
Figure A.23: Turn-on of L2J(1,8) in multi-jet events.
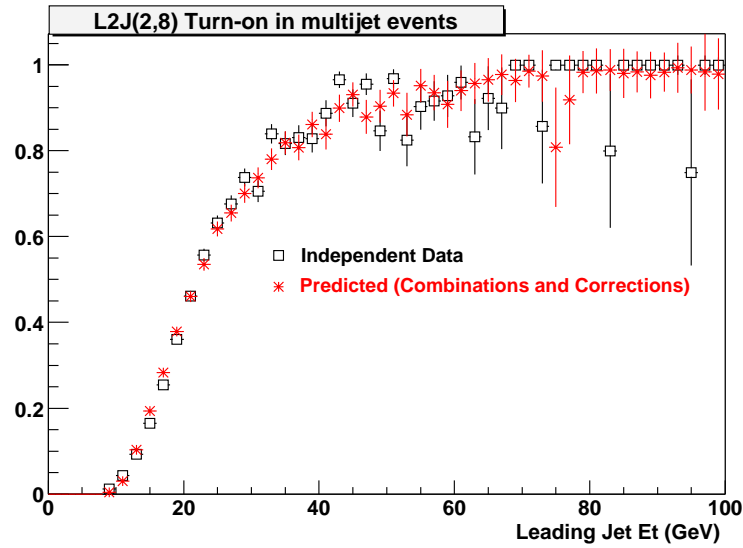
Figure A.24: Turn-on of L2J(2,8) in multi-jet events.
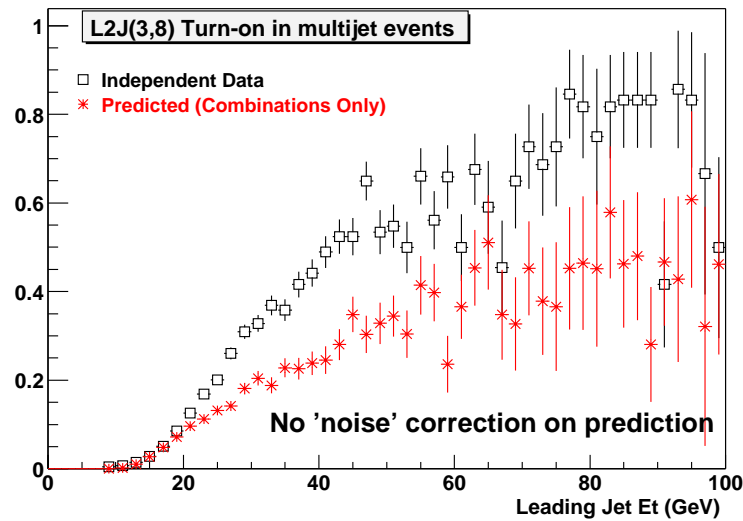
Figure A.25: Turn-on of L2J(3,8) in multi-jet events, with the double L2 jet from single offline jet probability not accounted for.
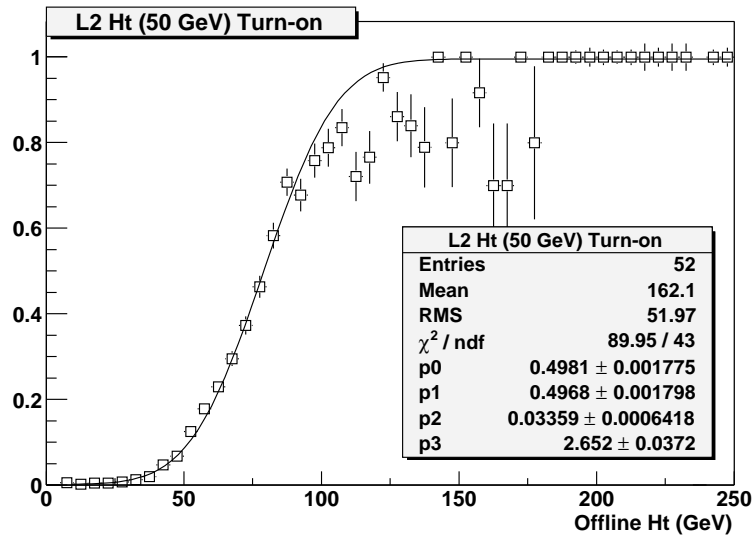
Figure A.26: Turn-on of L2H_t(50) in multi-jet events.

Appendix B

THE LEVEL 3 / DATA ACQUISITION SYSTEM

B.1 Introduction

The Data Acquisition System and Level3 Trigger at Fermilab's DØ Experiment assembles data from 63 VME crates each containing 1-20kB/event spread through 1-10 VME modules into complete 250kB events in one of 82 farm nodes at a 1kHz event rate. The events are run through fast reconstruction and 50Hz is selected and sent to tape storage. Up to 8 simultaneous runs are supported, allowing for flexible data taking, commissioning, and calibration. Near-real-time monitoring of each component is provided to many graphical displays of various types. Commodity hardware such as the CISCO 6509 Ethernet switch, open-source software such as ACE [2] and Linux, and the standard TCP/IP protocol are used throughout. The system has performed well since commissioning in July, 2002.

B.2 Hardware Components

The system is composed of Single-board Computers (SBCs), 100 Mb/s to Gb/s Ethernet converter switches (CISCO 2948Gs), a single main Ethernet switch (CISCO 6509), and computer farm nodes. The components and their connections are shown in Figure B.1. The SBCs send data from their VME crates over dual 100 Mb/s Ethernet connections to the 2948Gs which transfer the data to the main switch over Gb/s optical fibers. The farm nodes are connected to the main switch by 100 Mb/s Ethernet connections.

176
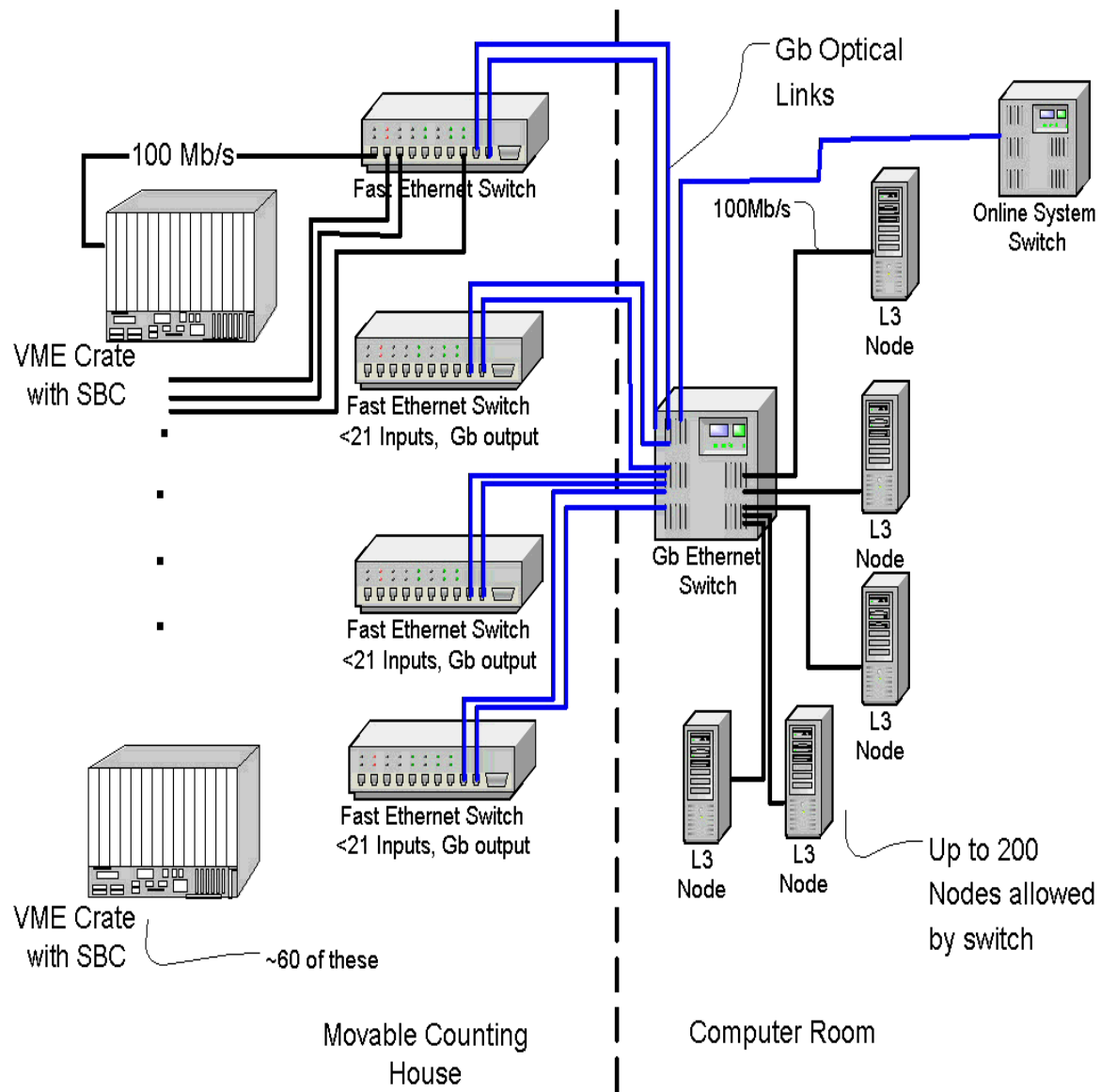
Figure B.1: Hardware components.

B.2.1 Single-Board Computers

The SBCs have two jobs in the system: they are readout controllers for the event data, and one hosts the Routing Master program which coordinates the others. The VMIC 7750 [43] SBC being used has a 933 MHz Pentium-III processor, 128 MB of RAM, 128 MB of flash ROM (used for booting), fast VME access, one PMC expansion slot, and two 100 Mb/s Ethernet interfaces. The VMIC SBCs use the Tundra Universe II [44] chip for the VME-PCI interface. The Universe II is compliant with 32 or 64-bit VME transfers, has a programmable DMA controller, offers flexible interrupt logic, and allows address pipelining, which is used in reading out some of the VME crates. The single PMC slot is occupied by a BVM [45] digital I/O module, used for coordinating the readout of VME modules over the VME J3 connector. Additional PMC slots can be added through an additional VME card, which may be used in the future to expand the Ethernet capabilities using Gb/s adapters.

The sole custom hardware component in the system is the extender board used to mechanically support the 6U commercial SBCs in the 9U readout crates, propagate signals to the panel at the front of the VME crate, and connect lines on the digital I/O card to the J3 backplane.

B.2.2 Ethernet Switches

Two types of Ethernet switches are used. A set of CISCO 2948G's act as concentrators, transferring packets from up to ten 100 Mb/s connections to Gb/s optical fibers. The number of 100 Mb/s connections is limited to ensure no network congestion occurs at this stage. The Gb fiber links are brought together in a central switch (the CISCO 6509) to which the farm nodes are also connected. The central switch can handle an effective bandwidth of 16 GB/s, which is more than sufficient for the peak dataflow of 500 MB/s. The CISCO 6509 can be configured with up to 9 modules. One module is being used to provide the 100 Mb/s Ethernet ports (for the farm nodes).

This module also has 112 MB of output buffer memory shared between the 48 ports. Another module contains the Gb/s optical fiber inputs (for accepting data from the 2948Gs).

B.2.3   Farm Nodes

Programs on the farm nodes build events from the data fragments received, reconstruct the events, and perform physics selection. The computers used are dual 1 GHz Pentium III rack-mounted machines with 1 GB of RAM, two 100 Mb/s Ethernet ports, and a small amount of local disk. There are currently 82 nodes, but the system is expandable to handle more than two times as many (limited only by the number of modules that can be added to the central switch).

B.3   Software Components

Custom software performs the logic used for reading out the VME crates, buffering and routing the event data, sending and receiving data across the Ethernet network, assembling the event fragments into complete events, and deciding whether the event should be kept or discarded.

Three main types of information need to be transmitted (and received): crate data, routing information, and node free-buffer information. In addition, monitoring information, remote commands, and configuration information also need to be passed, although at a lower rate and at lower priority. The TCP Ethernet protocol is used to transmit this data through the network with negligible loss or corruption, at high rate, with low latency, while consuming a minimum of CPU power cycles to send and receive. Although the UDP protocol was considered for broadcasting routing commands to the SBCs, TCP was chosen in the end for its predicable performance under heavy network load.

Instead of directly implementing our own TCP socket calls, the ACE [46] C++
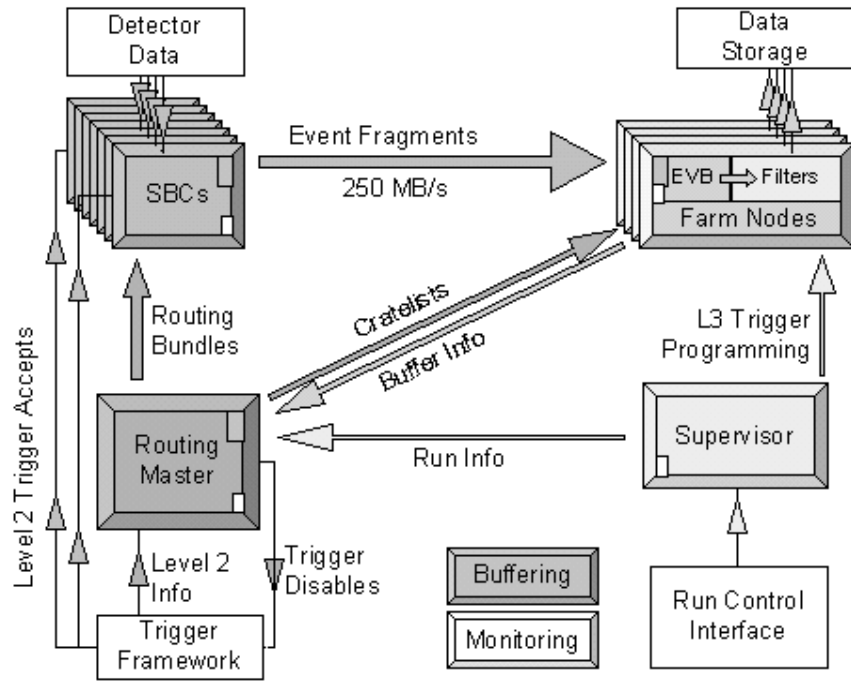
Figure B.2: Software components, showing dataflow and feedback.

Network Framework is used, which wraps this functionality while adding a negligible performance overhead. In addition, the ACE library offers many useful patterns such as thread management, thread-safe queues, timers, and de-multiplexing event handlers, through a platform independent, object-oriented interface. In particular, the *Task* construct, which combines an input and output queue with a worker thread, is the building block of all the routing and communication software.

B.3.1 Supervisor

The DØ experiment's data taking is controlled by shifters, who interact with run control software (COOR). The DAQ is designed to handle multiple simultaneous definitions of L1/L2/L3 triggers, each loosely referred to as a *run*. The Supervisor program provides the interface between COOR and the DAQ/L3 system. When a new run is configured, the Supervisor receives configuration information which specifies the

node resources to allocate to this run, the type of code to have the nodes execute, the VME crates that will be read out for each event in this run, and the trigger configuration, which contains a description of the requirements events have to satisfy to pass the L3 trigger. The Supervisor initializes the appropriate nodes and tells the Routing Master about this new run. The run can then be started and paused an unlimited number of times, and the Supervisor communicates the new state to the Routing Master so that it can decide whether to allow events to be sent to nodes in the run. When the run is stopped, the Supervisor instructs the nodes to return to their un-configured state, and downloads new programming to the Routing Master which no longer includes information about this run.

If an error occurs while the Supervisor is trying to change the state of the system, the change is aborted, and the whole system is restored to the last valid state. The error message is returned to the run-control system so that proper recovery steps can be taken. Programs which unexpectedly crash or are intentionally stopped and restarted are automatically reconfigured without shifter intervention.

B.3.2   Routing Master

The Routing Master program executes on an SBC in a special crate whose data contains information about which L1/L2 triggers passed. The RM's job is to synchronize and direct the sending of data from all the other read-out crates to one or more farm nodes for each event. The crate housing the RM also has special hardware connections back to the trigger framework, which the Routing Master can use to disable L1 triggers, to apply back-pressure when the L2 accept rate is too high.

The Supervisor configures the Routing Master with a set of runs that share farm resources through a strict set of rules. A run is a set of available nodes and triggers and a set of required crates that must be read out for each trigger. After each L2 trigger accept, the Routing Master reads data out of the special trigger crate and, for each run which overlaps with at least one of the passed L1/L2 triggers, it chooses one node

to be used for the event. This decision of which node to choose for each run is based on the number of available event *buffers* in each node. The Routing Master maintains an internal table of free buffers for each node, called the Node Free-buffer Map. Each time an event is routed to a node the corresponding entry in the Node Free-buffer Map is decremented by one. The entries are incremented periodically when the RM receives messages from the nodes, which indicate a number of event buffers that the node has freed. A node is chosen in a round-robin fashion from amongst the set of nodes with the most free buffers, to divide events amongst equally free nodes most evenly.

If too few free node buffers in a run are available (as determined from the Node Free-buffer Map), the Routing Master communicates over VME to other specialized cards in the crate it is housed to apply back-pressure to the L1 trigger system by disabling the triggers associated with that run. The triggers are re-enabled when a node in the run informs the Routing Master that free buffers are available again. The algorithm employed guarantees that the SBCs will not be instructed to send data to a node until the node is able to internally buffer the event immediately.

A set of VME crates to be read out for each run is generated by making the union of all the trigger lists associated with the L1/L2 triggers that passed for the event. The Routing Master creates a route command for each SBC which is in a VME crate in this set of crates. The route command is an integer representing the event number and a set of node indices indicating the nodes to which the SBC should send the data for this event. When all runs have been considered for an event, the route commands are added to routing bundles, which group together route commands for multiple events into a single message. This grouping is necessary to reduce the number of TCP messages sent out over Ethernet by the Routing Master, since there is a per-message overhead. A routing bundle containing a non-zero number of routing commands is sent out either after 250 ms or when it contains some maximum number (by default 10) of routing commands. A bundle is also sent early occasionally (.1%

of events) to add randomization, such that not all routing bundles are sent after the same event.

The Routing Master also sends information to each chosen node for each event: the event number, the crate list, and the fired L1/L2 triggers. The event number and crate list are required for the node to know when it has received data from all the crates needed for the event. The trigger list is needed for reporting to monitoring about which triggers were associated with an event, especially if the event was incomplete due to an error. Properly counting the number of events recorded for each trigger is crucial (for calculating the luminosity accurately).

In summary, the Routing Master polls a VME location connected to the Trigger Framework every 10ms and reads a list of Event Tags, containing a 16 bit L3 event number and a list of passed L2 trigger bits. For each Event Tag, it builds the list of runs which have at least one of its L2 trigger bits in the Event Tag. For each of these runs in the list:

- A Farm Node is chosen from the list of Farm Nodes in the run, based on which has the most free buffers in the Node Free-Buffer Map.

- For each L2 trigger bit in the Event Tag, the list of crates to be read out for that L2 trigger bit is added to the set of crates to be read out.

- The chosen Farm Node is sent a Cratelist containing the L3 event number, the L2 trigger bit list in the Event Tag, and the set of crates that will be read out.

- The index of the chosen Farm Node is added to a Route Tag to be sent to each crate in the set of crates to be read out. Each Route Tag is added the Routing Bundle for its SBC, which is soon sent to the SBC.

- If less than 16 Farm Node buffers are available, the trigger bits for the run are disabled. They are re-enabled when 24 Farm Node buffers in the run are free.

B.3.3 Read-Out Processes

The software running on the SBC in each standard read-out crate must read out
the data from other modules in the crate and send it to the nodes, as instructed by
routing commands from the Routing Master. The Linux operating system is used on
the SBCs for conformity with the majority of other systems in the experiment and to
take advantage of local expertise. Two main processes run on each board:

- The Interrupt Service Routine (ISR) is a kernel module which waits on a custom
  kernel interrupt initiated by a request for read-out communicated through the
  J3 VME connector to the BVM DIO module in the SBC's PMC slot. The ISR
  starts the data transfer (carried out through DMA) into a buffer in the SBC's
  memory. The buffers are created in kernel-space memory at boot-time through
  a specialized kernel module. The buffers are used in circular order, and two
  pointers keep track of the first and last buffer used, such that the data is never
  copied. This design, in particular the choice of putting functionality directly
  into the kernel itself, minimizes latency by avoiding the Linux scheduler.

- The Event Sender is a user-level process that receives routing information from
  the Routing Master and attempts to match it (by event number) with event
  data in buffers pointed to by the ISR. If the event number in a route command
  matches the event number in a data buffer, the Event Sender transmits the
  buffer's data to the specified node(s). If the event numbers do not match, a
  warning is issued, either the event data or the route command is discarded
  (based on which event number was behind), and a new match is attempted.
  This way, the Event Sender keeps itself synchronized even when routing or
  event data is occasionally missing or duplicated.

B.3.4 Farm Nodes

Three main types of processes execute on each farm node, also under the Linux operating system. These processes are managed by a Program Manager on each node, which can start the correct version of software, restart crashed processes, and stop processes.

- The Event Builder (EVB) combines event fragments, received from SBCs in the VME crates, into complete events. Event fragments are organized into complete events through the use of a map, keyed by event number. Data is read over Ethernet from the SBCs and buffered in a fixed-length input queue. Another thread processes event fragments in the input queue and places pointers to the fragments into the map. A map entry is marked as complete when it contains pointers to fragments from all the SBCs required, as specified by the crate-list sent to it by the Routing Master for that event. The fragments are then copied (concatenated) into a contiguous memory buffer provided by *IO Process* (see below) and deleted from the map. Map entries which are still incomplete after one second are erased, and the associated fragments are deleted. Incomplete events are carefully reported to the *Monitor Server* (see below). Whenever a map entry is freed, a message is sent to the Routing Master to inform it that space for another event is available, unless too many free buffers are already being advertised to the RM (see *Buffering* below).

- The IO Process takes complete events that have been built by the EVB and distributes them between the multiple L3 Filter processes (see below). Two L3 filter processes are needed to take advantage of both CPU's on each farm node.

- The L3 Filter processes reconstruct each physics event and decide if it passed or failed the programmed filter requirements. A passed event is sent over 100 Mb/s

Ethernet to a dedicated machine called the Collector, and failed events are discarded.

B.3.5 Monitor Server

The ability to retrieve detailed information from all software components in the system in near-real time (within a few tens of milliseconds) is crucial for debugging, troubleshooting, and understanding the performance of the system. The Monitor Server is software which runs on a dedicated commodity PC (similar to those used for receiving events), and coordinates the efficient gathering of this information.

All software components in the system (a total of over a hundred) connect to the Monitor Server as *clients*, and are able to respond to it with extensive statistics and other information about their status. The Monitor Server responds to queries from *displays*, such as a web-page scripts and graphical programs. The display's query can request any set of information from any set of components in the system. The Monitor Server then makes queries of the necessary client components, collects the information asynchronously as it is returned from the clients, and sends back a response containing the information initially requested by the display. The type of information gathered is very diverse, from event rates to the status of TCP connections to idle CPU time.

XML is used for all communications with the Monitor Server, as opposed to unstructured binary or textual data. The format is flexible enough to handle any kind of information transfer, safe and fast to parse in almost any language on almost any platform, and humanly readable for debugging purposes. The XERCES [47] parser was chosen for use in all C++ components for its speed, wide acceptance and support in the community, and ability to run on a wide variety of platforms.

The Monitor Server must be able to support a large number of displays operating worldwide, making simultaneous queries. However, there is some overhead in gathering information from a client, so care must be taken not to query clients too frequently, or the data-taking performance of the system would be degraded. There-
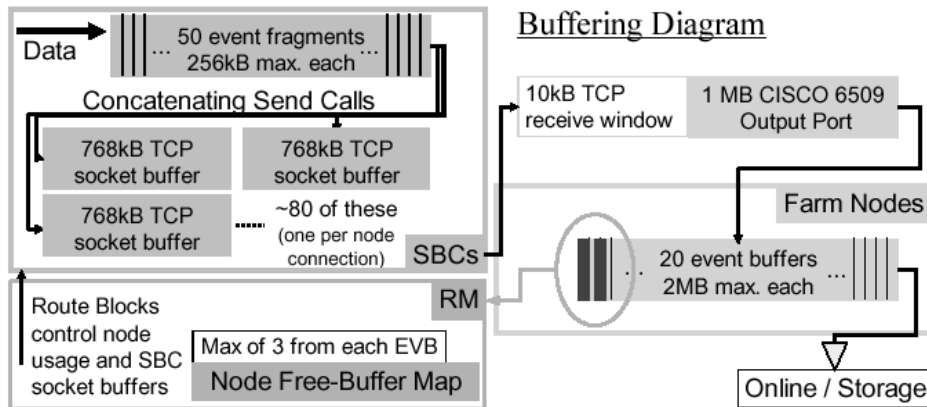
Figure B.3: Buffering controls.

fore, the Monitor Server caches information that it has gathered, so that no client will be bothered to report the same information more than once per second, no matter how frequently the server is queried.

B.4 Buffering

Operation of the DAQ at high rates requires that buffers be available to absorb latencies in the system or there will be unacceptable dead-time. In general, buffering exists almost everywhere data or other messages are transferred over Ethernet between machines and between threads or processes on the same machine.

Buffering in the system is adequate to sustain continuous data flow at full rate. The Routing Master controls data flow, based on free-buffer information fed back from the Farm Nodes, to take advantage of the distributed buffering throughout the system. If back-up occurs in or downstream of the Farm Nodes, the lack of advertised free-buffers from the Farm Nodes causes the Routing Master to disable triggers via the Trigger Framework interface before buffers in the system overflow.

Data read out over VME by an SBC is placed into one of 50 statically allocated, fixed-size kernel-memory buffers. These buffers are used in a circular order, with head

and tail pointers indicating the first and last used buffer, respectively.

Queues that can each hold approximately 100 routing commands exist in the RM to buffer the routing bundles waiting to be sent to each connected SBC. A corresponding queue in each SBC Event Sender program, of the same maximum size, buffers the routing messages on the receiving end as they are waiting to be matched with corresponding data read out from their VME crates.

Data sent from the SBCs to the farm nodes is buffered by the kernel-level TCP socket buffers. The connection's *send-buffer* size is set to 768 kB on the SBC, which provides about 100 event data fragment buffers on average (comparable to the routing message buffer). On the farm node, a queue with a maximum size of 40 fragments buffers event data fragments being received from all the SBCs. Fragments are buffered here until they are copied into the EVB's input queue, which has a maximum length of 700 fragments (enough to hold about 10 events). The EVB's event map can hold the pieces of at most 20 events at a time. IO Process then buffers these events until they are processed by one of the L3 filter processes. Finally, those events which pass the L3 trigger requirements are passed to a final output queue, where they wait to be sent to the Collector, a central machine that gathers events to be written to tape for offline analysis.

The number of event buffers the farm nodes *advertise* to the Routing Master (3) is not the full number they have available (20). The farm nodes each report a much smaller number of free buffers to ensure that the RM does not send out routing to an SBC that would cause it to fill up its TCP output socket buffer (causing the SBC to hang because it uses a blocking TCP send call). Since the maximum event fragment size is 256 kB, and the socket buffer size is 768 kB, the node only advertises 3 buffers at any one time to the RM. This is acceptable because, in total, the 82 farm nodes thus advertises 246 event buffers, more than in the SBCs.

The messages sent from the farm nodes to the Routing Master are buffered in a queue on each node which can hold a maximum of 100 messages. Once received in

the RM, the messages are processed immediately by the receiving thread to make the free node buffer usable again as soon as possible.

In summary, the following buffers and controls are used:

- Each SBC has 50 event fragment buffers in Linux kernel memory, where event fragments wait to be concatenated into TCP socket buffers.

- There is a large 768kB TCP socket buffer on each SBC $\rightarrow$ Farm Node connection. They can not fill up and block, since the maximum crate size is 256kB and the Routing Master only routes a maximum of 3 events to a Farm Node at a time. Ensuring that send calls will not block simplifies the event sending logic.

- The TCP_RTO_MIN compile-time Linux TCP parameter was changed from the 200ms default to 10ms in the Routing Master to assure no delays in Routing Bundle delivery due to occasional packet loss.

- Each output port on the CISCO 6509 has a 1MB buffer. The Farm Node TCP receive window size is kept at 10kB per SBC connection to not overflow the switch's output port buffer. (80 SBC connections x 10kB/connection < 1MB)

- The EVB has 20 event buffers where event fragments are built into complete events. Only 3 are advertised to the Routing Master, to not overflow the SBCs' TCP socket buffers.

- Events are held in 3 shared memory buffers waiting for the Filters to process them.

B.5   Control

B.5.1   Software Versioning and Distribution

An efficient and easily usable set of scripts exists for controlling the software that runs on each of the machines in the system. The currently active version of software on each machine can be specified from a central location. A version of software can be automatically distributed to machines if it does not already exist locally. The *rgang* [48] command is used to distribute the software efficiently, as well as to execute remote commands on multiple machines simultaneously.

B.5.2   Initialization

When programs involved with data-flow start up, they read a single cross-mounted (using NFS) XML file which defines the *topology* of the system. The file describes the DNS names of all the SBCs and farm nodes, assigns each node a unique index, and maps each SBC to its unique crate-id. These indices and crate-id's are used to identify each process in communicated messages at runtime. The topology file also identifies the location of the Supervisor, Routing Master, SBCs, and Monitor Server. Using this information, appropriate TCP connections are established between components in the system. The SBCs and farm nodes make the connection to the RM, since the RM's address is stable whereas addresses of other components, or even the number of them, can change over time. The farm nodes connect to the SBC's so that if a farm node crashes, it can reconnect to the SBC's when it is restarted.

B.5.3   Runtime Control

The Supervisor does the majority of the runtime control of the system, however it is often useful to be able to *manually* change the parameters of some set of processes while they are running, particularly while debugging or testing. To accomplish this, a configuration port is opened on the Routing Master and each SBC. A spe-

cial command-line program (executed either locally or remotely) can connect to this port and alter some parameters. For instance, in the Routing Master, the maximum number of routing commands in a routing group can be modified. In the SBCs, the compression settings (see below) can be changed, and an additional *stream* or file can be created for collecting some or all events. This last feature has been very useful for detector groups as they debug their systems.

B.6   Stability and Robustness

Several features were designed into the system to increase its stability and robustness, even in the face of unexpected mechanical, computational, or (most importantly) human errors. Every connection ensures that it is still viable by sending intermittent *pings* (either a short string or an empty TCP packet). If any connection is determined to be closed, either from the failure of a ping, or from the detection of a closed TCP socket, the remote machine is polled at a 10 second interval until the connection is re-established.

Most connections are also protected against hangs in the case that data can not be sent or received for some reason. There are 15 second timeouts on the sending or receiving of messages in many places, in case a remote machine is not behaving as expected. When receiving, checks are made that there is room to store the new data, and if there isn't, a warning is issued, and the oldest information is overwritten. (This can not occur unless a failure somewhere else in the system has caused the buffering logic to fail.) When sending, separate threads are often used to prevent an entire process from being blocked by one hung connection.

Many *sanity checks* are also included. For instance, the Event Builder periodically (at least every 5 seconds) re-communicates its availability to the Routing Master, in order to guarantee synchronization.

Also, the system is designed such that any set of machines or processes (with

the exception of the Supervisor) can be restarted without adversely affecting the rest of the system. Often, runs can continue unaffected, if they are not directly using the component being restarted, such as when a node or an SBC is restarted. If the Routing Master is restarted, all runs will stop until the needed components have reconnected to it, at which time all runs will resume automatically, without external intervention.

B.7 Ethernet Bandwidth Optimization

While the total bandwidth of the system is easily handled by the main CISCO 6509 switch, the maximum rate of the whole system is limited by the slowest crate. The bottleneck will most likely be the Ethernet transfer of data from a crate to nodes. Thus, care has been taken to optimize the Ethernet bandwidth out of the crates.

B.7.1 Dual Ethernet Interface Utilization

Each SBC has two 100 Mb/s Ethernet ports, and thus a maximum theoretical output bandwidth of about 24 MB/s. (Additional daughter cards can be added to the SBCs to give them GB/s Ethernet output.) Pooling the output bandwidth of both interfaces (which are on the same sub-net) to send data to a set of nodes is non-trivial, however. An Ethernet packet being routed by the operating system to any two nodes on the same sub-net will by default always choose the same hardware interface to use (unless explicit entries are added to the routing table for each node), regardless of which interface the connection was established on.

This default routing behavior can be avoided by using the BINDTODEVICE [49] socket option under Linux. Sockets are opened on two different ports, and each is bound to a separate hardware interface. Each node then makes a connection to each socket. The SBC alternates between the two ports on consecutive sends to different nodes and also on consecutive events to distribute the data most evenly between the

interfaces.

B.7.2 Dynamic Compression

Additional effective bandwidth can be achieved by compressing the data before it is sent, using a lossless compression algorithm, and then uncompressing it in the node. Two compression algorithms are available, and can be selected dynamically (see *runtime control* above). The gzip [50] algorithm offers the best compression factor, but uses a lot of CPU to compress the data. The lzo_1x [51] algorithm is amongst the fastest of compression algorithms, but does not achieve as great a compression ratio. To prevent the SBC from becoming CPU-limited, the fraction of events which are subjected to compression is dynamically adjusted to use less than 100% of the CPU power. In addition, in order not to waste CPU time, the fraction of compressed events is adjusted dynamically so that if the output bandwidth falls below a certain limit, the amount of compression is reduced.

B.8 Offline Testing

Since all components must be continuously operational, software changes need to be tested *offline* as much as possible, before being applied to the real system. The Routing Master, Event Sender, and Event Builder can be run in a *test* mode, which fully tests their code and simulates any inputs or hardware which do not exist. The Routing Master, for instance, can simulate incoming triggers at any desired rate. A full test, containing all these components, can be run on a single machine or multiple machines, and has been automated by simple scripts. In addition, the tests can be run under the Linux or Win32 platforms (or a combination), which helps to find odd bugs or race conditions that behave differently under the subtleties of each platform.

A test-crate which houses an SBC is used for testing new read-out hardware or software configurations. The data for the test crate can come from a variety of sources.

B.9 System Performance

The system has been running online continuously for over a year, and many tests have been performed. The performance of the full system has surpassed its design expectations.

The bandwidth output of the dual 100 Mb/s Ethernet interfaces of an SBC has been shown in the test crate to sustain greater than 22 MB/s sending to 4 nodes, each of which has a single 100 Mb/s Ethernet interface. The CPU consumption on the SBC was around 20%. Using the lzo_1x compression, average event data could be sent at over 30 MB/s, using nearly all of the SBC's CPU power.

High-rate tests have been run with up to 6 SBCs and 24 nodes. The Routing Master consumed about 30% CPU at 2 kHz. The CPU usage did not increase appreciably as more SBCs or nodes were added to the test (since the real work is in making the routing decisions themselves).

The Event Builder uses less than 10% of a CPU to build full events from 60 crates at 100 Hz. Recall that at the design rate of the system (1 kHz), each node only needs to process events at 20 Hz.

Downloading a new run is the most time-consuming control operation, since a large amount of information must be distributed to and parsed by each node. By sending commands in parallel, the Supervisor is able to program a farm of 24 nodes and the Routing Master in about 5 seconds total.

During normal running, there are routinely more than 50 monitor displays running in the control room and throughout the world. The Monitor Server typically has about 150 total clients connected, from which data is available. Gathering 10 kB of XML from a client takes about 10 ms. Responding to a query which demands gathering 300 kB of XML in total from all 90 clients takes about 50 ms, since information is gathered in parallel. The Monitor Server uses about 20% of its CPU power while serving about 500 kB/s of XML spread amongst many displays.

B.10   Conclusions

The software components rely upon high level programming languages; the Linux operating system; widely-used, open libraries; and standard networking protocols. The routing and buffering methods have performed reliably since commissioning and satisfy the current and future needs of the DØ experiment.

Using commercial hardware (SBCs, Ethernet switches, PC farm nodes) and widely-used software libraries (ACE, XERCES, etc.) helped to significantly reduce the amount of time necessary to design and implement a full Data Acquisition system. The SBCs are more than sufficient for simultaneously reading out data in the VME crates, receiving routing information, and sending the data over Ethernet to the chosen farm nodes. The Routing Master is able to generate and send routing messages to all the SBCs at beyond the required rate. The buffering strategies, monitoring system, software version control, and runtime control also work well. In addition to playing a critical role in our current experiment, this system is expandable to fulfill future DAQ needs at DØ and elsewhere.

VITA

Andrew Haas was born in Syracuse, NY, in 1976. He graduated from Fayetteville–Manlius High School in 1994. At Brown University he earned a B.S. in Physics in 1998. After receiving a M.S. in Physics from the University of Washington in 2000, he moved to Chicago, IL, to perform research at Fermilab. He has graduated with a Ph.D. in Physics from the University of Washington in 2004, and looks forward to continuing research in High–Energy Physics.